

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 07:40:07 ; Search time 9801 Seconds
(without alignments)
4454.459 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801

Sequence: 1 MALSMKPRADSDGFGSNL.....LAIQKQKQSSGTPAGQP 901

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool.h/US09920705/tunat_07042005_101813_20638/app_query.fasta_1.1095
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09920705_@CGN_1_1_5932_@tunat_07042005_101813_20638 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELACK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4801	100.0	2706	6	AX420420 Sequence
2	4786	99.7	2706	6	BT003029 Arabidops
3	4786	99.7	3498	8	AY094463 Arabidops
4	4656.5	97.0	3171	6	CQ804588 Sequence

5	4392	91.5	4593	6	AX420419	AX420419 Sequence
6	4392	91.5	102057	8	AC002335	AC002335 Arabidops
7	1436	29.9	801	6	AX420422	AX420422 Sequence
8	1365	28.4	3543	8	AK065222	AK065222 Oryza sat
9	1117.5	23.3	124083	8	AP005735	AP005735 Oryza sat
10	1117.5	23.3	154228	8	AC108755	AC108755 Oryza sat
11	1117.5	23.3	176553	8	AP006756	AP006756 Oryza sat
12	799	16.6	3455	8	AK100889	AK100889 Oryza sat
13	651.5	13.6	3715	6	AX420424	AX420424 Sequence
14	576	12.0	130461	8	AP004011	AP004011 Oryza sat
15	573.5	11.9	137519	8	ATT1P17	AL049730 Arabidops
16	573.5	11.9	196107	8	ATCHRIV34	AL161534 Arabidops
17	466	9.7	349	6	AX420423	AX420423 Sequence
18	374.5	7.8	116419	8	AC146743	AC146743 Medicago
19	332.5	6.9	6836	9	HSN303089	AJ303089 Homo sapi
20	331.5	6.9	6770	9	HSN303090	AJ303090 Homo sapi
21	302	6.3	2190	6	BD223778	BD223778 Human RNA
22	302	6.3	6057	9	HSN803154	AL831838 Homo sapi
23	302	6.3	168499	9	AC092037	AC092037 Homo sapi
24	302	6.3	174525	2	AC067763	AC067763 Homo sapi
25	297	6.2	1692	6	CQ736847	CQ736847 Sequence
26	287.5	6.0	2308	3	AT069434	AJ069434 Drosophila
27	287.5	6.0	3084	6	CQ577575	CQ577575 Sequence
28	287.5	6.0	3113	3	AF145664	AF145664 Drosophila
29	287.5	6.0	9744	6	CQ577574	CQ577574 Sequence
30	287.5	6.0	37983	2	AC020394	AC020394 Drosophila
31	287.5	6.0	86889	3	AC005448	AC005448 Drosophila
32	287.5	6.0	139410	3	AC007084	AC007084 Drosophila
33	287.5	6.0	254023	3	AE003839	AE003839 Drosophila
34	277	5.8	10839	10	AF156529	AF156529 Mus muscu
35	276.5	5.8	3384	9	AF368064	AF368064 Homo sapi
36	273	5.7	3192	6	BD159584	BD159584 Homo sapi
37	273	5.7	3192	6	AX882187	AX882187 Sequence
38	273	5.7	3192	9	AK022541	AK022541 Homo sapi
39	273	5.7	3312	9	AF368062	AF368062 Homo sapi
40	273	5.7	3423	9	AF368063	AF368063 Homo sapi
41	273	5.7	4192	5	BC070824	BC070824 Xenopus l
42	271.5	5.7	7815	9	HSN297259	AJ297259 Homo sapi
43	271.5	5.7	10718	10	AB055980	AB055980 Mus muscu
44	271.5	5.7	81137	9	AL355488	AL355488 Human RNA
45	271.5	5.7	158483	2	AC025987	AC025987 Homo sapi

ALIGNMENTS

RESULT 1	AX420420	2706 bp	DNA	linear	PAT 18-JUN-2002
LOCUS	AX420420	Sequence 2 from Patent WO0212518.			
DEFINITION	AX420420				
ACCESSION	AX420420.1	GI:21524576			
VERSION					
KEYWORDS					
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Amasino,R.M., Schomburg,F.M., Michaels,S.D. and Patton,D.				
AUTHORS	Patent: WO 0212518-A 2 14-FEB-2002;				
TITLE	WISCONSIN ALUMNI RESEARCH FOUNDATION (US)				
JOURNAL	Location/Qualifiers				
FEATURES	1..2706				
source	/organism="Arabidopsis thaliana"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:3702"				
	1..2706				
	/note="unnamed protein product"				
	/codon_start=1				
	/protein_id="CAD35485.1"				
	/db_xref="GI:21524577"				
	/translation="MALSMKPRADSDGFGSNLWVGSUTPTTSSDITELFGRYGDI				

DRITVYSRSGPFIYRHHVEBAVAKALOGANLNGSIOKIEVAPKPKCKSLWVGI
 GPNVSKDLEEFSEFKGIEDFRFLRKRKTAFLIDYEMDOLQKSMNGKMGSPR
 VDPISROAPKKEOMAGSYDNNGNHNKPOVPHYEPFKEDVQSKLWJGPETATQ
 CNDROILAHNAMILRGEIERHVSYPKRNPAIYERSEAEARQCEKGLRFPNPRKI
 MYSNDLEPBDODTFSFGMRKSTDMFNNDPSCVSPHSBTGIJGSRNPLRGNEBSY
 NGARINDVDEKPEWRRPSPAGTGLPSPGTGLPSPAGTGRPMKSNDSMTGSDP
 AQLVRESKTRRSDVDGFTPMGVDSRFGSGVAARPIGSPRSPDIHMGMLAKGDT
 PVCCARCPWPKGIGETKLEPVVNCSTARDNLMAKHYAVAIIGERIVFVREDREDFAS
 YTEFLYLSKDRAGVAKLDGDTLFLIPPSDITDYLQYTRBERLYGVVTKIPPAV
 PVTASYROESQSNPLHYMDODARSPANASISIPPRNTIRGAPBHLTAISKSVSE
 LRIRPMNAAPQAGVSLTPELATLTLIPATISQPAEBSHPMSGSTVSTAGSEL
 YNGEAPQAMRGQIVHDASNOSFQQYQNYTPAQGLPBPSPRYPASNSPMTGDM
 VHGMMQYOSQSVNMPOLSPENMHNNYSMTGSSNHPVPSQPMVOQYCEBAMPN
 YGPIPSYQOANFHVITNOALNPSPQFAMQBPADKALBEPONQRLRQPMISDGD
 QGTTGDEVDKNORYOSTLQFPAANLLQIQKQKQSSGTRPAGGP"
 46. .279
 /bound_molecy="RNA Binding Region"
 283. .522
 /bound_molecy="RNA Binding Region"
 610. .852
 /bound_molecy="RNA Binding Region"

ALIGNMENT Scores:

Pred. No.: 9.13e-184 Length: 2706
 Score: 4801.00 Matches: 901
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-920-705-3 (1-901) x AX420420 (1-2706)

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 DB 1 ATGGCGTATCTAATGAAGCATTCAGACCGATTCGCGTTTCCAGTCAACATCTT 60
 QY 21 TTPVALGIVSERLEUTHRPROGLIUTHRTHGLUSERASPLEUTHRGLIUEPHEGLIYAR 40
 DB 61 TGGGTGCGTAGCCTTAACGCGGAGACGACAGACTGACATCTGACGAGTTGTTGGAAGA 120
 QY 41 TYGIVASPILEAPRGILETHRVALTYRSESERARGLYPHEALAHETILEYTYR 60
 DB 121 TAGCGGATTTATTAATGATCAGCGTGATCTTCTCAAGGCGTTTGTATATACTAC 180
 QY 61 ARGHISVALGILUJUALAVALAIAALYSGIUALALEUGLNGIYVALASNEUBENGLY 80
 DB 181 AGACATGTGGAGGAGACAGTCCGACGCAAGGCGCTTCAAGAGCAAAATTGATGGA 240
 QY 81 SERGLNILEYSLILEGLIUTHRVALAARGPROIALYSPROCYELYSERLEUTHRVALGILY 100
 DB 241 AGTCAAAATTAAAGTTCGAATAACGACACCGCAAAACCTTGTAGAGTCTATAGCGTGGGT 300
 QY 101 GLIYILEGLYPROASNAVALSERLYSAPASPLEUGLUGLUPHESERILEYSLYYS 120
 DB 301 GGAAATGGCGCTTAATCTTCCAAAGATGACTCGAGAGAGAGATTGCGCAAGTTTGGGAAA 360
 QY 121 ILEGUASPHEARGPHELEUARGIUAARGLYSTHRVALAHEIIASPTYRTYRISLUMET 140
 DB 361 ATTCGAGAGATTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTAATTATAGATG 420
 QY 141 ASPASPALALEUGLALALYSESERMETASNGIYLYSPROMETGLYGLYSERPHLEUARG 160
 DB 421 GATGATGCTTTACAGGCTTAAGAGCATGATGAGAAAGCCTATGCGGTAGCTTTTGGGT 480
 QY 161 VALASPHELEUARGSERGINALAPROLYSLYSGIUGINTTPALAGIYSERTYRASPASN 180
 DB 481 GTTGAATTTCTCCGGTCAACAAGCGCAAAAAGAAACANTGGCGTGGCTTTTCATGATAC 540
 QY 181 ARGASNGIYASNEASNHISLYSPROGLINTYRPROHISSERTYRGUASPHELYSGIY 200
 DB 541 AGAAATGGCAATATGAAATCAATAACCGCAGATATCTCTCACTATATGAGACTTTAAAGGA 600

QY 201 ASPVALGILNPROSERLYSVALLEUTHRPIILEGLYPHEPROPROTHRALATHRGLNCYSASN 220
 DB 601 GATGTCACCAAGATGAAGGTTCTGAGATGGGTTCCCTCTCACTGATCAACATATGCAAT 660
 QY 221 ASRGILGILILEHNSANALAMETILEUEPHEGLIYGLIILEGLIARGVALIYSSER 240
 DB 661 GATGAGCAAAATCTGACAAATCGATGATCTTCTTGAGANTCGAAGGGGTAAAGT 720
 QY 241 TYRPROSERARGASNPHEALALEUVALGIUPHEARGSERIALAGLUJUALAARGGLNYS 260
 DB 721 TACCATCAAGAAATTTTGCACCTTGCTGAGTTTAGAGCGGGAGAGAGCTGCCAATGC 780
 QY 261 LYSGLIUGIYLEUGLIGLYARGLEUPHEASNPASNPALAGIILEYSLIMETTYRSEASN 280
 DB 781 AAGGAAGGCTTACAGGGAGGTTTATCAATATCTTGAACCAAAATTATGACTCAAAC 840
 QY 281 ASPGLIUEUPROPROGLIUGIINASPASPHRSEPHETYSERGLYMETLYASARGSERARG 300
 DB 841 GATGATGTCCTCCCTGAGACAGACGATCTAGTTTTCCTGATGATAACGGTCAAGG 900
 QY 301 THRASPMEUPHEASNPASNPASNPASNPASNPASNPASNPASNPASNPASNPASNPAS 320
 DB 901 ACAGATATGATCAATATATGATCTTCAATGATCTTCTCTCAATTCATGAGATTCCT 960
 QY 321 GLYSERMETARGPROLEUARGGLYTHRSENGIUAARGSERTYRANGIYALAGIUTHYRASN 340
 DB 961 GGGTCTTATGAGGCCCCCTCAGAGGTACGATAGAGGCTTCAATATATGAGTCAAT 1020
 QY 341 ASPVALVALGILYSGIUPROASNPTRPARARGPROSERALASANGIYTHRGLIYLEU 360
 DB 1021 GACGTTGTTGGTAAGAGACCAACCTGGAGAGGCGCATCTGAATATGAACTGGAATCTC 1080
 QY 361 PROSERPROTHRGLYPROGLIYILEUEUPROSERPROIALAGIINGIYTHRARGARGPROMET 380
 DB 1081 CCATCTCCAAACAGACCTCGAATCTCCATCTCTCCACAGAGTACAGAGCGCCCTATG 1140
 QY 381 ARGSEASNPROASNPSETRTPGLUGIYTYRASPPROIALAGIUEUVALARGIUSERLYS 400
 DB 1141 AGGTCAAAACCCGATTTCTGGAGAGATGATGATCTCTCAAGTTGGTCAAGAGAAAGTAAA 1200
 QY 401 ARGTHRARGARGARGIYSEVALASPGIYPHETHRPROMETGLYVALASPGIUAARGSER 420
 DB 1201 CGAACCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 QY 421 PHEGLIYARGIYSEVALAIAALARGPROIALARGGLYPROPROASNPSEASPHISILE 440
 DB 1261 TTGCTGAGAGTTAGTTGCTGCTAGACCTATCGTGGCCCCCTGATTCGATACATA 1320
 QY 441 TTPARGIYMETILEALALYSGIYGLYTHRPROVALCYSCYSAIARGCYSAIARGCYSAI 460
 DB 1321 TGGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 QY 461 GLIYISGLIYILEGLIUTHRLYLEUPROGLIUALVALASNCYSERIALAARGTHRASPLEU 480
 DB 1381 GGAAAGGAGATTGAATACTGAGAGTCCGAGAGTCAATGTTTCAACAGAACTGATTTG 1440
 QY 481 ASNEUASNEUVALYSHISTYRVALAVALAIAILEGLICYGILNILEVALPHEPHEVALPRO 500
 DB 1441 AATATGCTCCCTTAACCTTCCGCGCTTGCCATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
 QY 501 ASPARGIUGIUNASPHEALASERTYRTHGLUPHELEUARGTYRLEUSESERLYSASP 520
 DB 1501 GACAGGGAAGAAATTTTGGCTCTTACACTGAATTTCTCGGATACCTTGAAGCTCAAAAGAT 1560
 QY 521 ARGVALGILYVALAIALYSELEUASNPARGIYTHRTHLEUPHELEUVALPROPROSERASP 540
 DB 1561 CGGCGGGGTTGCCCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 QY 541 PHELEUTHRASPVALIUEUGIUNALTHRARGINGIUAARGIYGLYVALIYALLEUYS 560
 DB 1621 TTCTTAACGATGATCTTCAAGTGAACCTCAAGAACGCTATATGAGTTTGTCTCAAG 1680
 QY 561 LEUPROPROPROVALIUEUVALIUEUVALIUEUVALIUEUVALIUEUVALIUEUVALIUEU 580

```

Db      1681 TTAACCCCGCCGCGCTTCCTGTTACAGATCATATACAGACAGAAATCTCAGTCCATCT 1740
Qy      581 LeuHisTyrMetAspGlnAlaIarGAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
Db      1741 CTGCATTAATATGATCATAGCCCGGAGATTCACCTGCCAATCTGATCAGATTATATCTCT 1800
Qy      601 ProArgGluAsnTyrTrileuArgGlyAlaProGluHisLeuThrAlaLeuSerLysProSer 620
Db      1801 CCGAGGAAATATCAATTAGGGGTGACCAAGAACATTGACAGCTCTCTCAAAAACATCT 1860
Qy      621 ValSerGluProLeuArgGlyLeuProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
Db      1861 GTTAGGAGACCTCTCAGAACTACTAATATATGCGCCCTCAAGCTGGGTAGTTAACT 1920
Qy      641 ProGluLeuLeuAlaThrLeuAlaSerTrileuProAlaThrSerGlnProAlaAlaPro 660
Db      1921 CCGAGACTTTTAGCCACTCTGGCATCTATCTCCCTGCAACTCTCAACCTGCTGCCCT 1980
Qy      661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
Db      1981 GAGAGTCACCAACCTATGTCAGACCTTCAACAGTTGTTCCACAGCATCATCTCAAT 2040
Qy      681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
Db      2041 GGAAGCTTCAATGAGAAAGCAACCTCTCAAGCTTGAGAAAAGAGTCCACAAACAGTTCA 2100
Qy      701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
Db      2101 GATGCTCAATCAGTCAATTCACCAACATCGGAATCAGATCACCTCAGCTGGGCAACAT 2160
Qy      721 ProProProProSerArgTyrProProAlaSerAsnAsnProSerTyrThrSerGlyMet 740
Db      2161 CCTCTCTCTCTCTGCGTTACCTCCAGCTTCAACCAACCCCACTACCTACTGGAAG 2220
Qy      741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
Db      2221 GTTCAGGCAACATGCAATACCGAAGCAATCTGTTAACTGCTGCTGCTCCGTTA 2280
Qy      761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
Db      2281 CCAAAATATGCTCATATATATATTCATGATACACTCAGGGTTCGTTCAATCATCTCTG 2340
Qy      781 SerGlnProMetValGlnGlnTyrGlnProGlnAlaSerMetProAsnGlnAsnTyrGly 800
Db      2341 TCTCAGCCCATGCTCCAGCAATACCAACAGAAAGCGTCCATGCAAACTATGCT 2400
Qy      801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
Db      2401 CCAATTCAGATTATACAGAGCTAATTTTCATGCGGTAAACAACAATCAGGCACAGAAC 2460
Qy      821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGln 840
Db      2461 TTAAACCTCTCCCAATTTTCAAGCTGCAATGCAACCAAGATTAAGGAATTTTAAG 2520
Qy      841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
Db      2521 CCACAAACCAACAGCACTAGCATTTGACGCTATATCTCTGGGATGCTCAGGGTACAAAC 2580
Qy      861 AspGlyGlyValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
Db      2581 GATGGGAGAGTGCATAGATCAGATACAGATACAGTCACTACATCAATTTGCAAGAACT 2640
Qy      881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
Db      2641 CTCTCTCAATACAGAGAAACAGCAGCAACAGTCTTCAAGTACTCCGGCTGGAACAGGG 2700
Qy      901 Pro 901
Db      2701 CCT 2703

```

RESULT 2
BT003029

```

LOCUS      BT003029                2706 bp    mRNA       linear   PLN 16-JAN-2003
DEFINITION Arabidopsis thaliana At2g43410/Tt024.15 gene, complete cds.
ACCESSION  BT003029
VERSION    BT003029.1
KEYWORDS   FLI CDNA.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE  1 (bases 1 to 2706)
            Cheuk,R., Chen,H., Kim,C.J., Shim,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Becker,J.R.
            Arabidopsis ORF clones
            Unpublished
            2 (bases 1 to 2706)
            Cheuk,R., Chen,H., Kim,C.J., Shim,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Becker,J.R.
            Direct Submission
            Submitted (16-JAN-2003) Salk Institute Genomic Analysis Laboratory
            (SIGAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT    RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
            Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.
            The Salk, Stanford, PGSC (SSP) Consortium members constructed and
            sequenced the pUN1 (ORF) clones using the RAPL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Shim,P., Bowser,L., Chan,M.M., Chang,C.M.,
            Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yamada,K., Yu,G., Yuan,S., Davis,R.W., Theologis,A., and
            Becker,J.R.
            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Becker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES   Location/Qualifiers
            source          1..2706
                        /organism="Arabidopsis thaliana"
                        /mol_type="mRNA"
                        /db_xref="taxon:3702"
                        /chromosome="2"
                        /clone="U21966"
                        /ecotype="Columbia"
                        /note="This clone is in pUN1 51"
            CDS              1..2706
                        /note="unknown protein"
                        /codon_start=1
                        /product="At2g43410/Tt024.15"
                        /protein_id="AA023594.1"
                        /db_xref="GI:27764945"
                        /translation="MALSMKPRADDSGFGSNNIMVGLTETTESDITLPRGYDI
                        DRTIVYSSRGFAFIYRHVEBAVAKALQGANLIDVENDQIAKIEYAPAPRCSLWGI
                        GPNVSKDLSEFEFSFKGIEDFRFLERKRTAFIDYENDMDQIAKSNKMGSGFLR
                        VDFLSQAPKKEOMAGSYDNRNGNNHNFQVPHSYEDKGVOPSKVLMIGPPATQ
                        CNDQOILHNMILFGEIERVKSYPERNALVBFRAEERQCKEIQRLFNPRIKI

```

MYNDELPEQDDTSFYSGMKRSRTDMFNNDPSFVSSPHSNGI PGSMRPRGRNERSY
 NGAEYNDVVGKPMWRPSANGTGLI LSPSGPGLIPBPAGTRPRMNSNDSEGDIP
 AOIYREKRTPRDSDSGFTPMGVDES FGRGSAAPRIGPSPDSIDWGMTAKGT
 PVCCARCPYMKGIETLPEVNCSTADTLMKHAVALIGEIVFVREDREBPAS
 YTEFLRYSKDRAGVAKLDGDTLLFLVPSDEFLVDLQVTRBERLKI.PPAPV
 PVTAASYOESNPLAHYMDQADSPANASHLVPRENYRGAPEHLTAASKRVSSE
 LRITNNAPQAKGVSLETPELLATLASII.PATSOPAAPSHOPMSGSPVSTANOSGL
 YNGEAPSOAKRGPOTVHDASNOSFOOYGQYTPAGQLPPSPRYPSPASNPYTSGM
 VHGMOYOSOSVNMPLSP.LPMNHNYSYTOGSSNHPISQPMVOQIOPBAMPMON
 YGPIPSYQOANFHVTTNNOANLPSPOAPMQPAPKAKALEPQNALRLOPMISGDG
 QGTTDSVDKNORISTLOFPAANLLQIQKROQSSGTPAGGP

ORIGIN

Alignment Scores:

Pred. No.:	3.64e-183	Length:	2706
Score:	4786.00	Matches:	899
Percent Similarity:	99.89%	Conservative:	1
Best Local Similarity:	99.78%	Mismatches:	0
Query Match:	99.69%	Indels:	0
DB:	8	Gaps:	0

US-09-920-705-3 (1-901) x BT003029 (1-2706)

QY 1 MetAlaLeuSerMetLeuProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 Db 1 ATGGCGTTATCTATGAGCCATTGAGCCGATGATCCGGTTTCCAGTCAAAACATCTT 60
 QY 21 TrpValGlySerLeuThrProGluThrGlnSerAspLeuThrGluLeuPheGlyArg 40
 Db 61 TGGGTGGTGAGCTTACGCCGAGACGACAGACTGCAATCTGACCAAGTGGTTGGAGAA 120
 QY 41 TyrGlyAspIleAspArgIleThrValIlyrSerSerArgGlyPheAlaPheIleTyr 60
 Db 121 TACGGCGCATATGTATGAAATCACGGTGATTTCTTCAAGAGCGCTTCGTTATATATCTAC 180
 QY 61 ArgHisValGluGluValAlaAlaIlyPheGluAlaLeuGlnGlyValaAsnLeuAngly 80
 Db 181 AGACATGTGGAGAGAGAGAGCGAGCGCAAGAGGCTTTCAAGAGCAAAATTTGATGGA 240
 QY 81 SerGlnIleLeuIleGluTyrAlaArgProAlaIlyProCysLeuSerLeuThrValGly 100
 Db 241 AGTCAATTAAGTTCAGATACGACGACGACGACGACGACGACGACGACGACGACGACG 300
 QY 101 GlyIleGlyProAsnValSerIlyAspAspLeuGluGluIlyPheSerIlyAspGly 120
 Db 301 GGAATGGCCCTAATGTCTTCCAGAGATGACTGAGAGAGAGATTCAGCAAGTTTGAGAA 360
 QY 121 IleGluAspPheArgPheLeuArgGluArgIlyThrAlaPheIleAspTyrTyrGluMet 140
 Db 361 ATCGAGATTTTNGTCTTCTCAGAGACGACGACGACGACGACGACGACGACGACGACG 420
 QY 141 AspAspAlaLeuGlnAlaIlySerMetAsnGlyIlyrProMetGlyGlySerPheLeuArg 160
 Db 421 GATGATGCTTTACAGGCTAAGACATGATGAAAGCCATGGGTGATGCTTTTGCT 480
 QY 161 ValAspPheLeuArgSerGlnAlaProIlyIlyrGluGlnIlyThrAlaGlySerIlyAsp 180
 Db 481 GTTGATTTTCTCCGTCACAGCGCCCAAAAGAAACAATGGCGCTCTTACCATATAC 540
 QY 181 ArgAsnGlyAsnMetAsnHisIlyrProGlnTyrProHisSerTyrGluAspPheLeuGly 200
 Db 541 ABAATGGCAATATGATATATACCGCATATCTTCACTCATATGAAAGACTTTAAAGAA 600
 QY 201 AspValGlnProSerIlyValLeuThrIleGlyPheProProThrAlaThrGlnCysAsn 220
 Db 601 GATGTCCAGCCCAAGTAAAGGTCTGTGATTTGGTTCCTCCCTCACTGCTACCAATGCAAT 660
 QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValIlySer 240
 Db 661 GATGCAAAATTTCTCACAATGCGATGATCTTTGGTGAAGATCGAAGAGGTAAAGAT 720
 QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260

Db 721 TACCATCAAGAAATTTTGCACTTGAGATTAGAGCGCGAGCAAGCTCGCCAAATGC 780
 QY 261 IysGluGlyLeuGlnGlyValrGluPheAsnAsnProArgIleIlyMetTyrSerAsn 280
 Db 781 AAGGAAGGCTTACAGGGAGAGTTATCTCAATATCTTCAATATCAAAATTAATGACTCAAC 840
 QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetIlyAspSerArg 300
 Db 841 GATGATGCTCTCCGAGCAAGACATCTACTGTTTACTCTGATATGAACCGTCAAG 900
 QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
 Db 901 ACGATATGTTCAATATATGATCTTCAATTTGATATCTTCCATCTTACGGAATTCCT 960
 QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValGluTyrAsn 340
 Db 961 GGGTCAAGAGGCGCCCTCAGAGGTACAAATAGGGCTCATATATGCGTCAAGATATACAT 1020
 QY 341 AspValIleGlyIlyrGluProAsnThrArgArgProSerAlaAsnGlyThrGlyIleLeu 360
 Db 1021 GACGTTGTTGGTAAAGACCAAACTGAGAGGCGCATCTGCAAAATGGAATGAAATCTC 1080
 QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnIlyThrArgArgProMet 380
 Db 1081 CCATCTCAACAGACCTGGAATCTCTCCATCTCTGCAAGGATGAGGCGCTTATG 1140
 QY 381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerIly 400
 Db 1141 AGGTCAAAACCCGATCTTGGGAAGATATGATCTCTGCTGATGGTCAAGAAAGTAA 1200
 QY 401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
 Db 1201 GAAACCAAGAAAGATGATCGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATG 1260
 QY 421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
 Db 1261 TTTGGTGAAGGTTCAAGTTGCTGACATATCGGTGCGCCCTGATTTGTATACATA 1320
 QY 441 TrpArgGlyMetIleAlaIlyrGlyIlyrProValCysCysAlaArgCysValProMet 460
 Db 1321 TGGAGAGAAAGATTTGCAAGGGGTGAACTCCGCTGTTGTCTGCTGTTGTATGACCTATG 1380
 QY 461 GlyIysGlyIleGluTyrIlyrLeuProGluValValAsnCysSerIlyAlaArgThrAspLeu 480
 Db 1381 GGAAGGGAGATGAACCTTAACCTCCGAGGTGCTCAATTTGTTCAAGAAACCTGATTTG 1440
 QY 481 AsnMetLeuAlaIlyrHisIlyrValAlaValAlaIleGlyCysGluIleValPhePheValPro 500
 Db 1441 AATATGCTCGCTAAACATTAACGCGGTGCGATTTGATGATCGTTTTTTTCGTACCA 1500
 QY 501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerIlyAsp 520
 Db 1501 GACGAGAAAGATTTTGGCTTTACACGAAATTTCTCCGATACCTTTAGCTCAAAAGAT 1560
 QY 521 ArgAlaGlyValAlaIlyrLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
 Db 1561 CGGGCGGGTGTCCAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 QY 541 PheLeuThrAspValLeuGlnValThrArgGlnIlyArgLeuTyrGlyValValLeuIly 560
 Db 1621 TTTCTTAATGATATGATCTCCAGAGGACCGCTCAAGAACGCTATATGCTGTTGCTCAAG 1680
 QY 561 LeuProProProAlaValProValThrAlaSerTyrArgGluGlnIlySerIlyAsnPro 580
 Db 1681 TTACCCCGGACCGCTGCTGTTACAGATCAATACAGCAAGAAATCTCAGTCAATCTCT 1740
 QY 581 LeuHisIlyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
 Db 1741 CTGATTAATATGATCAAGCCGGGATTCACCGTCAAGATGATGATGATGATGATGATGAT 1800
 QY 601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerIlyrPro 620
 Db 1801 CTTAGGGAATAATTAATTAAGGGGTGCAACCAAGCAATTTGACAGCTGCTTCAAAACATCT 1860

QY 621 ValSerGluProLeuValGlyLeuProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
 Db 1861 GTTAGGAGACCTCTCGAATACCTTAATTAATGACGCCCTCAAGCTGGGCTTGAATTAACCT 1920
 QY 641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
 Db 1921 CCGGAGCTTTAGCCACTCTGGCATCTATTCCTCCGCAACTTCTCACTGCTGCCCCCT 1980
 QY 661 GluSerHisGlnProMetSerSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
 Db 1981 GAGAGACACCAACCTATGTCAGGACCTTCAACAGCTTGTTCACACACATCATGCTCAAT 2040
 QY 681 GlyLeuTyrAsnGlyValAlaProSerSerGlnAlaAlaThrValArgGlyProGlnThrValHis 700
 Db 2041 GGAAGCTGTAAGAAGAACGACCGTCTCAAGCTTGAAAAAGAGTCCAAACAGTTCAT 2100
 QY 701 AspAlaSerAsnGlnSerPheGlnGlnGlyValAsnGlnTyrThrProAlaGlyGlnLeu 720
 Db 2101 GATGCTCAATCATGATCTTCCACAAATACGAAATCATGATCATCTCACTGGGCAACTA 2160
 QY 721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
 Db 2161 CCTCCCTCTCTCGGCTTACCTCCAGCTTCAACACCCCACTACACTAGTGAATG 2220
 QY 741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
 Db 2221 GTTCATGAGCAACATGCAATACCAAGACCATCTGTTAACATGCTCAGCTGCTCGGTA 2280
 QY 761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
 Db 2281 CCAATATATCCAT 2340
 QY 781 SerGlnProMetValGlnGlnGlnTyrGlnProGlnAlaSerMetProAsnGlnAsnTyrGly 800
 Db 2341 TCTCAGCCCATGGTCCAGCAATATCCAAACCAAGAGCTCCATGCAACCAAACTATGCT 2400
 QY 801 ProIleProSerTyrGlnGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
 Db 2401 CCAATATATCCAT 2460
 QY 821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspValAsnLeuGln 840
 Db 2461 TTAAACCTCTTCCAT 2520
 QY 841 ProGlnAsnGlnAlaLeuAlaGlnGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
 Db 2521 CCACAAAAACCAAGCACTACGATTCAGCTATGATCTCTGGGATGCTCAGGGTACACA 2580
 QY 861 AspGlyValValAspValAsnGlnAlaGlyTyrGlnSerThrLeuGlnPheAlaAlaLeu 880
 Db 2581 GATGGGAGAGTGCATTAAGATACAGATACCACTCAACTCAATTTGACAGCAAACTCT 2640
 QY 881 LeuLeuGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 900
 Db 2641 CTTCTTCAATACAGCAAAACGCGACAGCAACAGCTTTCAGGATCTCCGCTGACAGGG 2700
 QY 901 Pro 901
 Db 2701 CCT 2703
 RESULT 3
 AY094463 3498 bp mRNA linear PLN 05-MAY-2002
 LOCUS Arabidopsis thaliana At2g43410/T1024.15 mRNA, complete cds.
 DEFINITION
 AY094463
 ACCESSION
 VERSION
 AY094463.1 GI:20453184
 KEYWORDS
 FLI CDNA
 SOURCE
 Arabidopsis thaliana (thale cress)
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
 AUTHORS
 1 (bases 1 to 3498)
 Shim, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Ban, J.,
 Bower, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamuya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
 Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
 Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.W.,
 Theologis, A., and Ecker, J.R.
 Arabidopsis cDNA clones
 Unpublished
 2 (bases 1 to 3498)
 Shim, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Ban, J.,
 Bower, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamuya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
 Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
 Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.W.,
 Theologis, A., and Ecker, J.R.
 Direct Submission
 Submitted (08-APR-2002) Salk Institute Genomic Analysis Laboratory
 (Signal), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA
 COMMENT
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN
 Arabidopsis Full-Length cDNA) : Seki, M., Narusaka, M., Ishida, J.,
 Hayashizaki, Y., and Shinzaki, K.
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Shim, P., Chen, H.,
 Cheuk, R., Kim, C.J., Meyers, M.C., Ban, J., Bower, L., Chang, E.,
 Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B.,
 Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
 Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A.,
 and Ecker, J.R.
 Shin, P., (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinzaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.
 FEATURES
 source
 CDS
 1..3498
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="2"
 /clone="RAFL09-47-P10(R21966)"
 /ecotype="Columbia"
 255..2960
 /note="unknown protein"
 /codon_start=1
 /product="At2g43410/T1024.15"
 /protein_id="AAM19833.1"
 /db_xref="GI:20453185"
 /translation="MALSKPEPRADDSGFGSNLNVGSLTPETTESDITLFGRYGDI
 DRIIVYSRGFAFIYRHYEBAVAAKALQGNLNGSOIKIEYARAPARKSLAWGDI
 GNVSKDLSEEFSEKLEIDREFLRERKTAIDYEMDDALQASNMKGKPGSGFLR
 VDFLASQPKKQOMAGSYDNRGNMNHPOYHSDYEDKDPQVSKULWIGCPRATO
 CNDDEIILNANILFEIERVYSYPERNLAIVFRSEARQCKEQLQRLPNNPIKI
 MTSNDELPEDDDTISFYSGMKSRKMDNRNDSPVSPHSTIPESMKPLRTNERSY
 NGAENVDVAGKEPNMRNSANGTILPSPGTGILPSPAQTRRPMRNSPDSWEGYD
 AOLVRESKRTKRGSDVDGFTPMGVDERSEFGSGSVARPIRGPDSDHLMRGMIAAGCT
 PYCARCVPMGKIGETKLPYVNGSARFDLNNLAGVAGIACEIVFVDPDEDFAS
 YTEPLRYLSSDRAGAVKLDGTTTLFLVPSGSLFDVLOVTRQELHLYGVKLPPPAV
 PRTASVROSGSNPLHYNDQARDSPANASHSYPRERYTIGAPHLTASPSVSE
 LTIENNAAPQAGVSILPELALATLILPATSOPAPAPESHQPSGSLVSTHQSNGL
 VNGEAPSQAMKRGQTVHADSNQSFQYQGYNTPTGQLPPEPSRYPASNNPNVYSGM
 VHGNQYOSGVNMPOLSPILPMPNNYSMTYQGSNNPVSQPMVQYQPEASMPNOM
 YGPIYSYOANHGVTTOANLNPSCQOANQPPADKANLEPQDQARLQPMISGDG
 GQTTGCVKDKQRVSTIQPANNLLQIQKRRQGSQSGTPAGQP"

```

misc_difference 2917
/notes="compared to genome"
/replace="a"

ORIGIN

Alignment Scores:
Pred. No.: 4,78e-183 Length: 3498
Score: 4786.00 Matches: 899
Percent Similarity: 99.89% Conservative: 1
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 8 Gaps: 0

US-09-920-705-3 (1-901) x AY094463 (1-3498)

QY 1 MetAlaLeuSerMetLysProPheAArgAlaAspAspSerGlyPheGlnSerAsnAlaLeu 20
Db 255 ATGGCGTTATCTATGAAAGCCATTCAAGCCGATGATTCGCTTCCAGTCAAACTTTCTT 314
QY 21 TrpValGlySerLeuThrProGluThrThcGluSerAspLeuThrGluLeuPheGlyArg 40
Db 315 TGGGTGGGTAGCCTTAACGCCGAGACGACAGCTCAGATTCGACCGAGTTGTTGGAGA 374
QY 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
Db 375 TACGGGGAATTTGATGAAATCAAGGTGATTTCTTCAAGAGCTTTCGTTTATATCTAC 434
QY 61 ArgHisValGluGluAlaValAlaAlaIleValGluAlaLeuGlnGlyValAsnLeuAsnGly 80
Db 435 AGACATGTGGAGAGACAGCTCGACGCAAGAGGCTCTTCAGAGCAAAATTTGAATGGA 494
QY 81 SerGlnIleLeuIleGluTyrAlaArgProAlaIleProCysIleSerLeuTrpValGly 100
Db 495 AGTCAATTAATTAAGATTCGAATTCGACGACCGCAAACTTTGAAGCTTATAGGTGGGT 554
QY 101 GlyIleGlyProAsnValSerIleAspAspLeuGluGluPheSerIlePheGlyIle 120
Db 555 GGAATGGGCTTATATGCTCCCAAGATGACCTGGAGAGAGATTCAGCAAGTTGGGAAA 614
QY 121 IleGluAspPheLeuArgPheLeuArgGluArgIleThrAlaPheIleAspTyrTyrGluMet 140
Db 615 ATCGAAGATTTTAAAGTTCTTCAGAGAACCCAGACAGCTTCAATGATTAATATGAGATG 674
QY 141 AspAspAlaLeuGlnAlaIleSerMetAsnGlyIleProMetGlyIleSerPheLeuArg 160
Db 675 GATGAAGCTTTACAGCTTAAGACATGAATGGAAGCCATAGGAGTGTGCTTTTGGCT 734
QY 161 ValAspPheLeuArgSerGlnAlaProIleIleValGluGlnTrpAlaGlySerTyrAspAsn 180
Db 735 GTTGATTTTCTCGGTCAAGGCGCCAAAGAAAAGCAATGGGCTGAGCTTTACGATTAAC 794
QY 181 ArgAsnGlyAsnMetAsnHisIleIleProGlnTyrProHisSerTyrGluAspPheLeuArg 200
Db 795 AGAAATGCAATATGATCATTAACCGGACGATCTCCACTCATATGAAGACTTTAAAGAA 854
QY 201 AspValGlnProSerIleValIleuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db 855 GATGTCACAGCCAAAGTAAGGTCTGTGAGATTGGGTTCCTCCCTACGCTTACCAATGCAAT 914
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValIleSer 240
Db 915 GATGAGCAAAATTTCTGCACAAATGCGATGATCTTTTGGTAGATGGAAGGGGTAAAGT 974
QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db 975 TACCATCATCAAGAAATTTTGCACTGTGGAGTTTGAAGCGGAGAGAACTGCCAAATGC 1034
QY 261 LysGluGlyLeuGluGlyArgLeuPheAsnAsnProAlaGlyIleValIleMetTyrSerAsn 280
Db 1035 AAGGAAGGCTTACAGAGGAGGTATTCTCAATATCTTAAGATCAAAATATATGACTCAAC 1094
QY 281 AspGluLeuProGluGluAspAspThrSerPheTyrSerGlyMetIleArgSerArg 300

```

```

Db 1095 GATGAGTTCCTCTCGAGCAGACGATATCTAGTTTATCTGTGATGAAACGCTCAAG 1154
QY 301 ThrAspMetPheAsnAsnAspProSerCysValIleSerProHisSerThrGlyIlePro 320
Db 1155 ACAGATATGTTCAATATATGATCTTTCATTTGATATCTTCCCTACCTACAGAAATTCCT 1214
QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValGluTyrAsn 340
Db 1215 GGGTCTATGAGGCCCTCCAGAGGTACAAATGAGCGTTATATATGTCAGAAATACAT 1274
QY 341 AspValValGlyIleValGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360
Db 1275 GACGTTGTTGGTAAGAGCCAACTGAGAGGCGCATCTGCAATGGAAGTGAATCTC 1334
QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
Db 1335 CCATCTCCAAACAGAGACTGGAATCTCCCACTCTCGCAGAAAGTACAGAGCGCCCTATG 1394
QY 381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerIle 400
Db 1395 AGGTCAAAACCCGATTTTGGGAAGATATATCTCTGCTCAGTTGCTCAGAGAAAGTAAA 1454
QY 401 ArgThrArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
Db 1455 CGAACCAAGAGATGATCATGTCAGCGTTTATCTCCAAATGGGTTCAGTAGAGGTCA 1514
QY 421 PheGlyArgGlySerValAlaIleArgProIleArgGlyProProAspSerAspHisIle 440
Db 1515 TTTGGTCAGAGTTCAAGTTGCTGCTAGACCTATCCGTGGGCCCTCGATTCTGATCACA 1574
QY 441 TrpArgGlyMetIleAlaIleValGlyIleThrProValCysCysValAspGlyValProMet 460
Db 1575 TCGAGAGGAATGATTTGCCAGAGGTGAGACTCCGCTGTTGTGCTGTGTGATCTATG 1634
QY 461 GlyIleGlyIleGluTrpIleLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
Db 1635 GGAAGGGGATTGAACCTAAACCTCGTAGAGTCTCATATTGTTAGCAAGACGATTTG 1694
QY 481 AsnMetLeuAlaIleValHisIleTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
Db 1695 AATATGCTCGGTAAACATTAACGCGGTGCCATTGGAGTGAATCGTTTTTTCGTACCA 1754
QY 501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrIleSerSerIleAsp 520
Db 1755 GACAGGGAGAAAGATTTTGGCTCTTACATGATTTTCCGGTACCTTACCTCAAAAGAT 1814
QY 521 ArgAlaGlyValAlaIleValLeuAspAspGlyThrThrIleuPheLeuValProProSerAsp 540
Db 1815 CGGCGGGGTGTGCCAAATTAATGATGATGATCAAACTTATTTGGTGCTCCATCATGAT 1874
QY 541 PheLeuThrAspValIleGluIleValThrArgGlnGluArgLeuTyrGlyValIleLeuIle 560
Db 1875 TTCTTAATGATGATCTCCAAAGTAAACCGTCAAGAACGGCTATATGAGTGTGTTCTCAAG 1934
QY 561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580
Db 1935 TTACCCCGGCAAGCGGTCTCTGTTAAGCATCATTAACGACAAAGAACTCAGTCAATCCT 1994
QY 581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
Db 1995 CTGATTAATATGATCAAGCCCGGATTCACCTGCCAAATGCTAAGTCAAGTTATATCTCT 2054
QY 601 ProArgGluAsnTyrIleArgGlyValProGluHisIleuThrAlaIleSerIleProSer 620
Db 2055 CTRHGGAAATTAATTAAGGTGTCACAGAACATTTGACAGCTCTTCAAAACCATCT 2114
QY 621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
Db 2115 GTTAGGAGGCTCTCAGAAATACCTAATATGACAGGCGCTCAAGCTGGGTATGTTAACT 2174
QY 641 ProGluLeuLeuAlaThrIleuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
Db 2175 CCGAGACTTTTACCACTCGGATCTATTTCTCCGCAACCTTCTCAACCTGTGCCCT 2234

```

```

QY 661 GUserHiGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
DB 2235 GAGAGTCACCAACCTATGTCAGACCTTCAACGGTGTTCACACACATCATGCTCAAT 2294
QY 681 G1LeuTyrAsnG1yG1uAlaProSerGlnAlaTrpLysArgG1yProGlnThrValHis 700
DB 2295 GGAAGCTTACAAAGAGACACCGCTTCAAGCTTGAAAAAGAGCTCACAAACAGTTCTAT 2354
QY 701 AAspAlaSerAsnGlnSerPheGlnGlnTyrG1yAsnGlnTyrThrProAlaG1yGlnLeu 720
DB 2355 GATGCGTCAATACGATTCATCCAAATACGAAATCAGTACACTCAGCTGGGCAACTA 2414
QY 721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
DB 2415 CCGCTCCCTCCCTTGGGTTACCTCCAGCTTCAAAACCCCAACTACATCTATGGAAG 2474
QY 741 ValHisG1yAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
DB 2475 GTCCATGCGCAATGCAATACAGAGCCCAATCTGTTAACATGCTTACGCTGTCCGTTA 2534
QY 761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnG1ySerSerAsnHisProVal 780
DB 2535 CCAAAATATGCGCTCATATATATTATTCATGTAACACTCAGGGTTCGCAATCATCTCTGT 2594
QY 781 SerGlnProMetValGlnGlnTyrGlnProGlnAlaSerMetProAsnGlnAsnTyrG1y 800
DB 2595 TCTCAGCCCATGCTCCAGCAATACCAACAGAACGCTTCATGCAACCAAACTATGAT 2654
QY 801 ProIleProSerTyrGlnGlnAlaAsnPheHisG1yValThrThrAsnGlnAlaGlnAsn 820
DB 2655 CCAATTCAGATTATACAGACGTAATTTTCATGGGTTACACAAATCAGGACAGAAC 2714
QY 821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGln 840
DB 2715 TTTAAACCTTCCCAATTTCAAGCTGCCATGCAACCAACAGACATTAAGGCAATTTAGAG 2774
QY 841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetHisSerG1yAspG1yGlnG1yTyrThr 860
DB 2775 CCACAAACCCAGCACTAGCATTTGACGCTATGATCTCTGGGATGCTCAGGGTTACAACA 2834
QY 861 AAspG1yG1uValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAsnLeu 880
DB 2835 GATGGGAGAGTCATTAAGATTCAGAGATTCAGACTCACTCAATTTGACAGCAAACTT 2894
QY 881 LeuLeuGln11leGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 900
DB 2895 CTTCTCCAGATACAGCAAAACGCGACAGCAACATCTTCAAGTACTCGGCTGACAGGGG 2954
QY 901 Pro 901
DB 2955 CCT 2957

RESULT 4
CQ0804588 3171 bp DNA linear PAT 10-MAY-2004
LOCUS CQ0804588
DEFINITION Sequence 999 from Patent WO2004035798.
ACCESSION CQ0804588
VERSION CQ0804588.1 GI:47110928
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1 Inze, D., de Veylder, L. and Vliethe, K.
Identification of novel e2f target genes and use thereof
Patent: WO 2004035798-A 999 29-APR-2004;
CropDesign N.V. (BB)
FEATURES
location/Qualifiers
1..3171
/organism="Arabidopsis thaliana"

```

```

ORIGIN .. /mol_type="unassigned DNA"
/db_xref="taxon:3702"

Alignment Scores:
Pred. No.: 6,6e-178 Length: 3171
Score: 4656,50 Matches: 892
Percent Similarity: 85,04% Conservative: 6
Best Local Similarity: 84,47% Mismatches: 3
Query Match: 96,99% Indels: 155
DB: 6 Gaps: 3

US-09-920-705-3 (1-901) x CQ0804588 (1-3171)

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
DB 1 ATGGCGTTATCTATATAGCAATTCAGACCGGATGATTCGGTTTCCAGCAAACTATCT 60
QY 21 TrpValG1ySerLeuThrProGlnThrGlnSerAspLeuThrG1uLeuPheG1yArg 40
DB 61 TGGGTGGTAGCTTAACCGCGAGACGACAGATCAATCTGACGAGTTGTTGAGACA 120
QY 41 TyrG1yAsp11eAspArg11eThrValTyrSerSerArgG1yPheAlaPhe11eTyr 60
DB 121 TACGGGATATGATAGATACACGGGTGATTTCTCACAGAGCTTGGCGTTATATATAC 180
QY 61 ArgHisValG1uGlnAlaValAlaAlaLysG1uAlaLeuGlnG1yAlaAsnLeuAsnG1y 80
DB 181 AGACATGTGAGAGAGACAGTCCAGCCAAAGAGGCTCTTCAAGAGCAAAATTTGAAATGA 240
QY 81 Ser
DB 241 AGCAATGAGCTTCGAAAAAATATACATCAACATGTTTACTCTCGTTTATGCCAAGA 300
QY 81
DB 301 GGTTAATCCATGTCGTTATCTGCAAACTGTATCCGCGTTTACGAGCTGATTTCA 360
QY 81
DB 361 GGTTCCTCAGATGTTAATCTCAACAGCTTGAATGATATGTTGGTACTCCATC 420
QY 82
DB 421 GAAATTCGATATAGAAATTTTGGAGTTTAAAGTTCACTTCTGCAAAATCAAA 480
QY 87
DB 481 CCGCGTGAGAAATTTGTCTATGGGGTACGTAGATCAATATCATGATCGCAACA 540
QY 87
DB 541 GAAATTCGCAACCGTATCGTCTTAATCTAGATGCTGAGACATATCCATATCAG 600
QY 87
DB 601 TGTGAGAGAGTTTGGGCTTGTGTTGCTCGATGCTTCAGAAACAATTAAGTGGTTG 660
QY 87
DB 661 TTAAGCGCGCTAATCTACCAATCAGAAACGTTTATTCGAGAACAATGTTGTTCTC 720
QY 88 ---AlaArgPro---AlaLysProCysLysSerLeuTrpValG1yG1y11leG1yProAsn 105
DB 721 AATTCACATCCCTTAGCAAAACCTTGTAAAGATCTATAGGTGGGTGGAATCGGCTTAAT 780
QY 106 ValSerLysAspAspLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 125
DB 781 GTCTCCAAAGATGACCTTGAGGAGAGATTCACCAAGTTGGGAAAAATTCAGGATTTTAA 840
QY 126 PheLeuArgG1uArgLysThrAlaPhe11eAspTyrTyrG1uMetAspAspAlaLeuGln 145
DB 841 TTTCTCAGGAACGCAAGACAGCTTCAATGATTTATATGAGATGATGATGCTTTACAG 900

```

QY 146 AlalysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArg 165
 DB 901 GCTAAGACATGATGAAAGCCTTAGGGGAGTCTTTGGCGGTGATTTTCTCCGG 960
 QY 166 SerGlnAlaProLysGlyGlnTrpAlaGlySerTyraAspAsnArgAsnMet 185
 DB 961 TCACAAGCGCCAAAGAAACAAATGGCGCTTACGATACAAATGGCAATATG 1020
 QY 186 AsnHisLysProGlnTyrProHisSerTyraGluAspPheLysGlyAspValGlnProSer 205
 DB 1021 AATATTAACCGCAGATCTCCTCACTCATATGAAACCTTTAAGAGATGTCCAGCCAACT 1080
 QY 206 LysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlnIleLeu 225
 DB 1081 AAGTTCGTGGATTTGGGTTCCCTCACTACGCTACAAATGATGACAAATTTCTG 1140
 QY 226 HisAsnAlaMetIleLeuPheGlyGlnIleGluArgValLysSerTyraProSerArgAsn 245
 DB 1141 CACAAATCGATGATACCTTTGGTGAATGAGAGGGGTAAAGTTACCATCAAGGAAT 1200
 QY 246 PheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlyLeuGln 265
 DB 1201 TTTCGACTTGTGAGTTAGAGGCGGAGAGAGCTCCCAATGCAAGAGAGCCCTACAG 1260
 QY 266 GlyArgLeuPheAsnAsnProArgIleLysIleMetTyraAsnAspGluLeuProPro 285
 DB 1261 GGGGGGTTATTCAATAATCTTAGAATCAAAATATATGATCAAAACATGAGTTCCTCT 1320
 QY 286 GluGluAspAspThrSerPheTyraSerGlyMetLysArgSerTyraAspMetPheAsn 305
 DB 1321 GAGCAAGACGATACGTATGTTTACTGTGATGAAAACGGTCAAGAGCAAGATATGTTCAAT 1380
 QY 306 AsnAspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgPro 325
 DB 1381 AATGATCCTTCAATTTGATTTCTCTCTCATTTCTACGAAATTCCTGGGCTATAGAGCC 1440
 QY 326 LeuArgGlyThrAsnGluArgSerTyraAsnGlyAlaGluTyraAsnAspValGlyLys 345
 DB 1441 CTCAGAGGATCAAAATGAGGCTTCAATTAATGATGAGCAATACATGACCTTTGGTAAAG 1500
 QY 346 GluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGly 365
 DB 1501 GAGCCAACTGAGAGAGGCCATTCGCAAAATGAACTGGAATCTCCCATCTCCACAGGA 1560
 QY 366 ProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAsp 385
 DB 1561 CCTGGAATCTCCCATCTCTCTGCACAAGGATCGAGGCGCCCTATGAGTCAAAACCCCAT 1620
 QY 386 SerTrpGluGlyTyraAspProAlaGlnLeuValArgGlySerLysArgThrArgArgAsp 405
 DB 1621 TCTTGGGAAGATATATCTCTGCTGAGTGCAGAGAAATGAACGAACCGAAGAGAT 1680
 QY 406 GlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
 DB 1681 GGATGAGTGAAGGTTTACTCTCAATGGGTGCTCGATGAGAGGTCTTTTGGTGAAGTTC 1740
 QY 426 ValAlaAlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIle 445
 DB 1741 GTTGTGCTAGACCATATCCGTGCGCCCTGATTCGATCAATATGAGAGAGATGATT 1800
 QY 446 AlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGlu 465
 DB 1801 GCCAAGGAGGAACTCCCTCTGTGTGTGCTCTGTGTGATCAATATGGGAAAGGGGATTAA 1860
 QY 466 ThrLysLeuProGluValAlaAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
 DB 1861 ACTAAACGCTGAGAGTCTGATTCATGTTCAAGCAAAACGATTTGAAATGCTCGCTAA 1920
 QY 486 HisTyraAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
 DB 1921 CATTAGCGCGTGTGCAATGATGATCGTTTTCGTACCAAGACAGAGGAAGAGAT 1980

QY 506 PheAlaSerTyraThrGluPheLeuArgTyraLeuSerSerTyraAspArgAlaGlyValAla 525
 DB 1981 TTGGCTCTTACACATGAATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGTGTGCC 2040
 QY 526 LysLeuAspAspGlyThrThrIleuPheLeuValProProSerAspPheLeuThrAspVal 545
 DB 2041 AATTTAGATGAAGGTACCACTTATCTTGGTGCTGCTCATCAAGATTTCTTAATCTATGTA 2100
 QY 546 LeuGlnValThrArgGlnGluArgLeuTyraGlyValValLeuLysLeuProProProAla 565
 DB 2101 CTCCAAGTACCCGCTCAAGAAAGCGCTATATGAGTGTGTTCTCAATTCACCCCGCACCC 2160
 QY 566 ValProValThrAlaSerTyraArgGlnGluSerGlnSerAsnProLeuHisTyraAsp 585
 DB 2161 GTTCTGTAGACATATACAGCAAAATCTCAGTCAATCTCTCATATATATGAT 2220
 QY 586 GlnAlaArgSerProAlaAsnAlaSerHisSerLeuTyraProProAlaGluAsnTyra 605
 DB 2221 CAAGCCCGGATTCACCTCCCAATCTAGTCAAGTTATATCTCTTAGGGAATATAC 2280
 QY 606 IleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluProLeu 625
 DB 2281 ATTAGGGGTGCACAGAACATTTGACAGCTGCTCAAAACATCTGTTAGGAGCTCTC 2340
 QY 626 ArgIleProAsnAsnAlaAlaProGluAlaGlyValSerLeuThrProGluLeuLeuAla 645
 DB 2341 AGAATACCTAATATGACGCGCTCAAGCTGGGTAGTTAACTCCGAGACTTTTAGCC 2400
 QY 646 ThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGluPro 665
 DB 2401 ACTCGCATCTATCTCCCTGCAACTTCTCAACTGCTGCGCTGAGAGTCAACAACT 2460
 QY 666 MetSerGlyProSerThrValAlaSerThrAlaHisGlnSerAsnGlyLeuTyraAsnGly 685
 DB 2461 ATGTAGGACCTTCAACAGTTGTTTCAACAGCAATCAATGACATGATGATCAATGGA 2520
 QY 686 GluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGln 705
 DB 2521 GAGAGCCGCTCAAGCTTGAAGAAAGAGTCCAAACAACTTCATGATGAGTCAAAACAG 2580
 QY 706 SerPheGlnGlnTyraGlyAsnGlnTyraProAlaGlyGlnLeuProProProSer 725
 DB 2581 TCATTTCCAAATACGGAATCAATGATCACTCAGCTGAGGCAATCACTCTCTCTCTG 2640
 QY 726 ArgTyraProProAlaSerAsnAsnProAsnTyraThrSerGlyMetValHisGlyAsnMet 745
 DB 2641 CTTTACCTTCAAGCTTCAAAACAAACCCCACTACATGATGGAATGATGATGCAATG 2700
 QY 746 GlnTyrglnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHis 765
 DB 2701 CAATACCAAGCCCAATCTGTTAATCATGCTCAAGCTGCTCGTTACCAATATAGCCCAT 2760
 QY 766 AsnAsnTyraSerMetTyraThrGlnGlySerSerAsnHisProValSerGlnProMetVal 785
 DB 2761 AATATTAATTCATGATCACTCAGGGTTCGTCAATATCATCTGTTCTCAGCCCATAGTCT 2820
 QY 786 GlnGlnTyrglnProGluAlaSerMetProAsnGlnAsnTyraGlyProIleProSerTyra 805
 DB 2821 CAGCAATACCAACAGAGAGCTCATGCCAAACCAAACTATGATTCATTCAGATTA 2880
 QY 806 GlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGln 825
 DB 2881 CAGCAAGCTAATTTTCAATGCGTATACCAACAAATCAAGCACAGAACTTAAACCTTCCAA 2940
 QY 826 PheGlnAlaAlaMetGlnProProAlaAspLysValAsnLeuGluProGlnAsnGlnAla 845
 DB 2941 TTTCACGCTCCATGCAACCAAGAGATGAAGCAAAATTTTAAGCAACAAACCAAGCA 3000
 QY 846 LeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAsp 865
 DB 3001 CTACGATTCGAGCTATGATCTCTGGGAGATGATCAGAGTATCAACAGATGGGAGTCTGAT 3060
 QY 866 LysAsnGlnArgTyrglnSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGln 885

Db 3061 ABAATACAGAGATACCACTACATTTGCGAACAACCTTCTCCAGATACAG 3120
QY 886 GlnLysGlnGlnGlnSerSerGlyThrProAlaGlyGlnGlyPro 901
Db 3121 CAGAAACAGACAGACAGTCTTCCAGGTACTCCGCTGACAGGGGCT 3168
RESULT 5
AX420419 4593 bp DNA linear PAT 18-JUN-2002
LOCUS Sequence 1 from Patent WO0212518.
DEFINITION AX420419
ACCESSION AX420419
VERSION AX420419.1 GI:21524575
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1
AUTHORS Masino, R.M., Schomburg, F.M., Michaels, S.D. and Patton, D.
TITLE Floral induction gene
JOURNAL Patent: WO 0212518-A 1 14-FEB-2002;
WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES
source 1..4593
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
ORIGIN
Alignment Scores:
Pred. No.: 3,81e-167 Length: 4593
Score: 4392.00 Matches: 899
Percent Similarity: 58.72% Conservative: 0
Best Local Similarity: 58.72% Mismatches: 2
Query Match: 91.48% Indels: 631
DB: 6 Gaps: 4
US-09-920-705-3 (1-901) x AX420419 (1-4593)
QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAenAenLeu 20
Db 1 ATGGCGTTATCTTGAAGCATTCAAGCGGATGATTCGGTTTCCAGTCAACATCTT 60
QY 21 TTPValGlySerLeuThrProGlnThrGlnSerAspLeuThrGlnLeuPheGlyArg 40
Db 61 TGGGTGGTAGCTTAACGCCGAGACGACAGATCTGACCGAGTTGTTGGAAGA 120
QY 41 TYRGLYAspLLeuAspArgLLeuThrValTYRSerSerArgGlyPheAlaPheLeuTYR 60
Db 121 TACGGGATATATGATGAATCAACGGGTATCTTCAAGAGCTTTCGTTATATATCTAC 180
QY 61 ArgHisValGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
Db 181 AGCATGTGATGA 240
QY 81 SerGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 90
Db 241 AGTCAATTTAATGATCAATTCAGCAGCCGGTTTGTCTTATCTATATCTTGTGTTG 300
QY 90 ----- 90
Db 301 TCTAATCTTGATGCTCTTTTGTCAAGATTATCTCTTTTGGAAATCATAGTCAGGT 360
QY 90 ----- 90
Db 361 TCACAAAACCTTGAATGCTGTTAGTCCAAAATTTCTGTGATCTGTTTTTTT 420
QY 90 ----- 90
Db 421 TTCCTATGATCAATCAAAAGTCGAAACCTAGTTTTTTTCTATATACGTCGTAGCTT 480

QY 90 ----- 90
Db 481 AAGGCGAAACCTGATCCGATCGAAAGCTTTTCTCAAAATTACTTGGTTATATGCACT 540
QY 90 ----- 90
Db 541 CCGGCAAGCCAAACACAGAGAGCTTCGAAAATTTGATGTTAAAGCATATATACTC 600
QY 90 ----- 90
Db 601 TTAGCAATGAGCTCTGCAAAAAGATACATCAACAGATTTACTCTGTTATGCGAA 660
QY 90 ----- 90
Db 661 GAAAGTTTATCCGATTTGCTTTATCTGCAAACTTGATCCGCTTACAGCTGATT 720
QY 90 ----- 90
Db 721 TCAGTTGCTTCAAGATGTTAATCTCAGACCTTGAGTGAATGATTTGGTACTCA 780
QY 90 ----- 90
Db 781 CTCGAATTCGAATGTGTAATAATTTGAAAGAGCAACTCTATCATCGGCCAAACAG 840
QY 90 ----- 90
Db 841 TAAGAATTTGAGATTAAAGTTCAAGTTCTCGCAAAAATCAAAACCGGTGAGAAAT 900
QY 90 ----- 90
Db 901 TTGTCTATGCGGTACGTATGATATCAATACTGCACTGCGAAACGAAATTTGGCAA 960
QY 90 ----- 90
Db 961 CCGTATCGTCTTAATCTTATGATGCTGCGACATATCCATATCAATGATGAGGAGTTT 1020
QY 90 ----- 90
Db 1021 GGGCTGTTTGGCTGATGCTTCAAGAAACAAATTACGTGTTGTTAAGCGGCTAAC 1080
QY 90 ----- 90
Db 1081 TCTACCAATCAGAAACGCTTATTTGGAAGAACATGTTTCTTCAATTCATCCCTTA 1140
QY 90 ----- 90
Db 1141 CGTACATCTGGGCTTTCCATTGATGTTCTTAGAAGTTGACTTCTTCAACCATTC 1200
QY 90 ----- 90
Db 1201 TGGGATGTGACAGTTGACGTGACAGACAAATATTCATTTGACGACGACTCAAATTT 1260
QY 90 ----- 90
Db 1261 CAATGCTTCCGGCTTGAATTAATGAGAGTTCAGGAGTTGTTATTTGGGAC 1320
QY 90 ----- 90
Db 1321 TCAGCAATGATTAAGACAGAAATGTTTGCAGAAAGTAAATGCTTCCGCTTGG 1380
QY 90 ----- 90
Db 1381 GGGACTACATCTTCTCTACAGACAAATCTGATGTTTAGCAAGAGAACTATTACA 1440
QY 90 ----- 90
Db 1441 CTGATTTATAGAGAAAGAGAGATGTTTCTCTTGGCGCAATTTTATTTGTAGAA 1500
QY 90 ----- 90
Db 1501 AGCAATGATATGAATGATGTCGTAGTGTAGTGAATTAATTAAGTATGTTGTGTTT 1560
QY 90 ----- 90

Db	1561	AGTTTCCTGATGTTGATGCTTATTCCTGGCAACCTATCTGGGTAGTATCGCCTCTT	1620	Db	2640	GGTACAAATGAGCGCTTCATATATAGTGACAAATACATGACGTTGGTAAAGACCA	2699
Qy	91	-----AlaLysProCysLysSerLeuTyrValGlyGlyLeuGly	103	Qy	348	AsnTyrArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly	367
Db	1621	ATCGACCTTTCTTCTGTTGGACGCAAAACCTTGTAAGTCTATAGGTGGGTGGGATCGGC	1680	Db	2700	AACTGAGGAGGCCATCTGCAATGAACTGGAAATCTCCATCTCCAAAGACCTGGA	2759
Qy	104	ProAsnValSerLysAspAspLeuGluGluLuphSerLysPheGlyLysIleGluAsp	123	Qy	368	IleLeuProSerProAlaGlnGlyThrArgArgProMetLysSerAsnProAspSerTyr	387
Db	1681	CCTATATGCTCCAAAGATGACCTGGAGGAAAGTTCCAGCAAGTTGGAAAAATCGACGAT	1740	Db	2760	ATCTCTCCATCTCTGCAACAAAGTACGAGCGCCCTTATGAGGTCAAAACCCGATCTTGG	2819
Qy	124	PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla	143	Qy	388	GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgAspAspGlySer	407
Db	1741	TTTAGGTTTCTCAAGAACCGACAGACGCTTCATGATTAATATGATGATGATGATGCT	1800	Db	2820	GAAAGATATGATCTGCTCAGTTGTCAGAAAGTAAACCAACAGAGAGATGATCA	2879
Qy	144	LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe	163	Qy	408	ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla	427
Db	1801	TTTACAGCTTAAGACATGAAATGCAAGCCCTATGAGGTGATGCTTTTTCGCTGATTTT	1860	Db	2880	GTGAGCGGTTTACTCCAAATGGGTGTCATGATGAGAGTCAATTTGGTCGAGTTGCT	2939
Qy	164	LeuArgSerGlnAlaProLysLys-----	171	Qy	428	AlaArgProIleArgGlyProProAspSerAspHisIleTyrArgGlyMetIleAlaLys	447
Db	1861	CTCGGTCACAGCGCCAAAAAGTAAGCACTTGTGGCATTTGATTTTACTTTTGA	1920	Db	2940	GCTAAGCCTATCCGCGCCCTGATTTCTGATCAATATGAGAGAAATGATGGCCAG	2999
Qy	171	-----	171	Qy	448	GlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLys	467
Db	1921	AAAGCTCCAGTAACAATTTGTTAGTTCAATATTTGGCTCAAACTGATAGGCGTGAG	1980	Db	3000	GGTGAACTCCGCTGCTGT	3059
Qy	171	-----	171	Qy	468	Leu-----	468
Db	1981	CTGTGCTTGCCCTAGAGACAGATTTACTGCTGCTCTATTTCATTGTAGATAGAGCT	2040	Db	3060	CTGATGACTAATTTCTAGACCTTAAACCTCTAGTGTCTTCTTTCAGAGCATTT	3119
Qy	172	-----GluGlnTyrAlaGly	176	Qy	468	-----	468
Db	2041	CAACTTCTTAAGTCTGAATCAAGTTACCTTTGTGTATCTTCAAGAAACAATGGCTGCG	2100	Db	3120	ATATATTTCCATTTCATTCTCGATGGAAGTACATTAATTAATAGTACATTTTATT	3179
Qy	177	SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln-----	190	Qy	468	-----	468
Db	2101	TCTTACATTAACGAATGCAATGCAATGCAATCAATAAACCGA-GGTTAGTCTGAATGTG	2159	Db	3180	TTACTATTACTGTTTGTAGTTCTGAGATGCTTGATTTTCATGTGTGATTCATTTTG	3239
Qy	190	-----	190	Qy	469	-----ProGluVal	471
Db	2160	AAAGTATGCTCTGTTACTAGTATGATATGATAGTTACTGGTTTGGAGTTTGTATTA	2219	Db	3240	GCATTCGCCCTCAATTAAGCTTTGTTTTTTTAAATTAATGATTTATAGGCTGAGGT	3299
Qy	191	-----TyrProHisSerTyrGlyAspPheLysGlyAspValGlnProSerLysVal	207	Qy	471	IleAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrValAlaValAla	491
Db	2220	TTCTTACAGATCTCTACTCATATGAAGCTTTAAAGAGATGTCAGCAAGTAAAGT	2279	Db	3300	CGTCAATTTGTTCAAGCAAGACTGATTTGAATATCTGCTGAACATTAAGCCCTTGCCAT	3359
Qy	208	LeuTyrPileGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn	227	Qy	491	eglyCysGluIleValAlaPhePheValProAspArgGluGluAspPheAlaSerTyrThrGly	511
Db	2280	CTGTGATTTGGGTTCCCTCTACTCTACTCATATGCAATGATGATGCAAAATTTCTGACAT	2339	Db	3360	TGATATGAGATCTGTTTTTTTTCGTACCAAGGGAAGAAATTTGGCTTACACTGA	3419
Qy	228	AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla	247	Qy	511	uPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyTh	531
Db	2340	GGCATATATCTCTTGTGTAGATCGAGAGGTAAAGTTAAAGTTCAACCAAGAAATTTTGC	2399	Db	3420	ATTTCTCGGTACTTATGCTCAAAAGATCGGGGGGTGGCCAAATTAATGATATGATG	3479
Qy	248	LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlyGlyLeuGlnGlyArg	267	Qy	531	rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGly	551
Db	2400	CTTTGAGATTTTAAAGCGGAGAGAGGCTCCCAATGCAAGAAAGGCTTACAGGGGAGG	2459	Db	3480	AACTTATTTCTGTGTGCTTCATCAATTTCTTAATGATATCTCAAGTACAGCCCTGA	3539
Qy	268	LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln	287	Qy	551	ngluArgLeuTyrGlyValValLeuLysLeuProProAlaValProValThrAlaSer	571
Db	2460	TTATTCATATATCTTAGATCAAAATTTATGATCTCAACGATGATTTGCTTCTGAGCA	2519	Db	3540	AGAAAGGCTATATATGAGGTGTTCTCAAGTTAACCCCGCAGCGCTTCTGTTACAGCATC	3599
Qy	288	AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspAsp	307	Qy	571	rTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAlaAspGlnAlaArgAspSer	591
Db	2520	GACGATACTAGTTTAACTGTGATGAACCGTCAAGACAGATATGTCATAATATGAT	2579	Db	3600	ATACAGCAAGATCTCAATCCATCTCTGCAATTAATGATCAAGAGCCGGGATTCAC	3659
Qy	308	ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg	327	Qy	591	AlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgValAlaProGly	611
Db	2580	CCTTATTTGATATCTTCTCTCATATCTGGAATTTCTGAGTCTATATAGGCCCTCTCA	2639	Db	3660	TGCCAATGTAGTACACAGTTTATATCTCTGAGGAAATTAATTAAGGGGTCACACAG	3719
Qy	328	GlyThrAsnGluArgSerTyrAsnGlyValGluTyrAsnAspValAlaGlyLysGluPro	347	Qy	611	uHisLeuThrAlaAlaSerLysProSerValSerGluProLeuArgGlyIleProAsnAsnAl	631
Db				Db	3720	ACATTTGACGCTGCTTCAAAACCATCTGTGTAGCGAGCTCTCAAGAAATCAATTAATATGC	3779

QY 631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuAlaThrLeuAlaSerIleLeu 651
 DB 3780 AACGGCTCAAGCTGGGGTAGTTAACTCCGAGCTTTTAAAGCCACTTAACTTCT 3839
 QY 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerThr 671
 DB 3840 CCTCGAATTCTCAACCTGCTGCCCTAGAGTCCCAACCTATCTCAGACCTTCAAC 3899
 QY 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyraGlnGlyValAlaProSerGlnAl 691
 DB 3900 ACTGTTGTTCCACAGCAGCATCAGTCCAAATGAGCTGTCAAATGAGGAAGCAACCGCTCAAC 3959
 QY 691 aThrPylAspGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnIlePro 711
 DB 3960 TTGGAAAAGAGGTCACAAACAGTTTATATGCGTAAATTCAGTCCATCCAAACAAATACG 4019
 QY 711 yAsnGlnIleThrProAlaGlyValLeuProProProProSerArgTyProProAlaSer 731
 DB 4020 AAATCAGTACACTCCAGCTGGGCACTACTCTCTCTGCGCTTACCTCCACTCAGCTTC 4079
 QY 731 rAsnAsnProAsnIleThrSerGlyMetValHisGlyAsnMetGlnTyrglnSerGlnSe 751
 DB 4080 AAACAACCCCACTACACTAGTGAATGTCATGTCACATGCAACATGCAATACAGAGCCAAATC 4139
 QY 751 rValAsnMetProGlnLeuSerProleuProAsnMetProHisAsnAsnIleSerMetTy 771
 DB 4140 TGTTAACAGTCCCTCAGCTGCTCTCCGTTACCAATATGCTCCATATATATATTCATGTA 4199
 QY 771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnIleTyrglnProG 791
 DB 4200 CACTCGAGGTTGCTCAATCATCTCTTCTCTCAGCCCATGCTCCACCAATACCAACACAGA 4259
 QY 791 uAlaSerMetProAsnGlnAsnIleTyrglyProIleProSerTyrglnGlnAlaAsnDheH 811
 DB 4260 AACGTCATCAGCCAAACCAAACTATGTCCAATTCACAGTATACAGCAACGTAATTTTCA 4319
 QY 811 eGlyValIleThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetG 831
 DB 4320 TGGCGTAAACAACAAATCAGGCACAGAACTTAAACCTTCCCAATTTTCAAGCTGCCATGCA 4379
 QY 831 nProProAlaAspLeuAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
 DB 4380 ACCACACAGAGATTAAGCAATTTAGAGCCACAAACCAAGCACTACGATTCAGAGCTTAT 4439
 QY 851 tIleSerGlyAspGlyValGlnIleThrThrAspGlyValAspLeuAsnGlnAlaArgTyrg 871
 DB 4440 GATCTCTGGGAGTGTCAAGGATCAACAAGATGGGAGGTGATTAAGATCAGAGATACCA 4499
 QY 871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGlnGlnIleGlnGlnIleGln 891
 DB 4500 GTCAACACATCAATTTGAGCAAAACCTTCTTCCAGATTAACAGCAAAACAGCAGCAACA 4559
 QY 891 nSerSerGlyThrProAlaGlyGlnIlePro 901
 DB 4560 GTCTTCAAGTACTCCGCTGACAGGGGCTT 4590
 RESULT 6
 LOCUS AC002335/c 102057 bp DNA linear PLN 11-MAR-2002
 DEFINITION Arabidopsis thaliana chromosome 2 clone T1024 map C1C10F02,
 complete sequence.
 ACCESSION AC002335
 VERSION AC002335.3 GI:20196857
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 102057)
 AUTHORS Rounaley,S.D., Techudy,M.M., Lin,X., Ketchum,K.A., Crosby,M.L.,
 Brandon,R.C., Spriggs,T.A., Mason,T.M., Kerlavage,A.R., Adams,M.D.,

JOURNAL Somerville,C.R. and Venter,J.C.
 REFERENCE 2 Unpublished
 AUTHORS Lin,X.
 JOURNAL Title
 TITLE Direct Submission
 REFERENCE 3 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 JOURNAL Title
 TITLE Direct Submission
 REFERENCE 3 Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, cdtowm@tigr.org
 COMMENT On Apr 18, 2002 this sequence version replaced gi:598343.
 FEATURES
 source location/Qualifiers
 1..102057
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C10F02"
 /clone="T1024"
 1..16126
 /note="overlap with BAC clone F14B2
 (AC004450:74158..90283)."
 complement(334..355)
 /rpt_family="AT-rich"
 complement(3113..4397)
 /gene="At2g43270"
 /note="synonym: T1024.1"
 complement(join(<3113..3691,3734..3924,4004..>4397))
 /gene="At2g43270"
 complement(join(3113..3691,3734..3924,4004..4397))
 /gene="At2g43270"
 /codon_start=1
 /product="putative myb-related transcription factor"
 /protein_id="AAB64309.2"
 /db_xref="GI:20196861"
 /translation="MIVYPPDLLEIFLGLPIKSLIRFQVSKQMSILSKSPARR
 LNVKKKELIAYGDRTELGFBGEESIQMYTLHCDIDATRPISLTBGLVCTPAPGMIV
 LNPSTRORRPPSCSNHNSDRIRFQFDELYLSPGNWMEGFRDKNFSYKVR
 MCFSPVEKCEVLDVETGESEINPPNDIDVGRKSRFGILADLTLEPHNFSVSTC
 CTSAMQNVNLDRLAITKTNGAEMELEISLMDYDKELISKYSITSLASITLHNWE
 SSWPYLVYLSKGINVPCDNHRLPKYNTETNEICLSDISVISPYLEIDVPLRLDS
 GHDHDKIRISSCGLFSGHPBAGLFKEITTRFVPAITLVGYQYRSLKPK"
 5148..5767
 /gene="At2g43280"
 /note="synonym: T1024.2; predicted by geneScan"
 join(<5148..5447,5552..>5767)
 /gene="At2g43280"
 join(5148..5447,5552..5767)
 /gene="At2g43280"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAB64334.1"
 /db_xref="GI:2289005"
 /translation="MPYDDYSRRLGFVNRVMSCRSEKDGRIARRFGCKKCHCVSI
 RKGFSVRRPSTREGCKAMTHVRSKGVITKVEVHNHPLVSVFRAHPTLDE
 KDKRIQELTIELRNKRLCAAYKEQIDAFKIVEBSNQIAKKVENVVNNLKEFEHLE
 HALLRSKDAI"
 complement(5982..6953)
 /gene="At2g43290"
 /note="synonym: T1024.3; supported by full length cDNA:
 Ceres:31535"
 complement(<5982..>6953)
 /gene="At2g43290"
 complement(6214..6861)
 /gene="At2g43290"
 /codon_start=1
 /product="putative calcium binding protein"
 /protein_id="AAB64310.2"
 /db_xref="GI:20196862"
 /translation="WVRIFLTVLNINSFLSLVPPKRLRTLPFLSWFDTKLTHKNSPPSP

```

repeat_region      complement (21849.. 21900)
                    /rpt_family="AT_rich"
gene               22675..22510
                    /gene="At2g43340"
                    /note="synonym: T1024.8"
                    join(<22675..22164,23243..23304,23325..23510)
mrna               /gene="At2g43340"
                    join(22675..23164,23243..23304,23325..23498)
                    /gene="At2g43340"
                    /codon_start=1
                    /product="unknown protein"
                    /protein_id="AAB64331.1"
                    /db_xref="GI:2289002"
                    /translation="MCDSPFLSRGCSYEADVDSDSEISTSPSSCSYSESEEREINNG
                    FGVOSKOTKKLEKKSSVTLLEGVNAVAVNDI.KPKSLTDDLEBKCVDFGRF
                    FNEEELPELCNTLPALPELCYSMSQKPTDQDHHHSSSPKKSVLDSVPVLSWKI
                    SSPGNDPVDYKRLKFMADVACTMRFGPDQPKAESWEEDDETRQFAGVLSLMRI
                    TCLKTIVVELFAHVTICIRLMI"
gene               /complement (23549..25505)
                    /gene="At2g43350"
                    /note="synonym: T1024.9; supported by full length cDNA:
                    Ceres:30737"
                    /complement (join(23549..23749,24086..24133,24215..24382,
                    24465..24883,24673..24734,24830..24906,25278..>25505))
                    /gene="At2g43350"
                    /complement (join(24101..24133,24215..24382,24465..24583,
                    24673..24734,24830..24906,25278..25439))
                    /gene="At2g43350"
                    /codon_start=1
                    /product="putative glutathione peroxidase"
                    /protein_id="AAB64335.1"
                    /db_xref="GI:2289006"
                    /translation="MPRSSRNQNRATSKIKFLLFGVAFVFLYLRPSSPTVEOS
                    STSINIVKDIIEGDVLSLKEFGKVLIVNVASKGLTGNVKEKMLLYAKKTOGF
                    ELIAPGQPGSOEBSGMEIKETVNICFKAEPIPKITVNGKNCPLYNPLKEQNG
                    GLFGAIIWNFAKFLINDRQGVNDVRYPTTSPLEIEDIKYKLASA"
                    /complement (25700..28116)
                    /gene="At2g43360"
                    /note="synonym: T1024.10; supported by full length cDNA:
                    Ceres:42038"

Alignment Scores:
Pred. No.:         1,03e-165          Length:      102057
Score:             4392.00            Matches:      899
Percent:           58.72%             Conservative: 0
Best Local Similarity: 58.72%          Mismatches:   2
Query Match:       91.48%             Indels:       631
DB:                8                  Gaps:         4

US-09-920-705-3 (1-901) x ACO02335 (1-102057)

QY      1 MetLaleuSerMetLpPProPheArgAlaAspAspSerGlyPheGlnSerAenAenLeu 20
Db      45895 ATGGGGTATCTAAGAACCAATTCAGACGACGATGATTCGCGTTTCCAGTCAACATCTT 45833
QY      21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
Db      45835 TGGGTCCGTACCTTAAGCCCGAGACACAGACAGTCAATCTGACCAAGTGTGGAGAA 45776
QY      41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaIleIleTyr 60
Db      45775 TACGGCATATGTATGATATCACGGTATTTCTTTCACGAGGCTTTCGTTTATATCTAC 45716
QY      61 ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAenLeuAngly 80
Db      45715 AGACATGTGAGAGAGAGAGTCGACGCCAAGAGAGCTTTCAGAGACAAATTTGAATGGA 45656
QY      81 SerGlnIleLeuIleGluTyrAlaAcArgPro----- 90
Db      45655 AGTCAATTTAAGATTCGAATTCGACGACGCGTTTGTCTTATCTATATCTTTCGTTTCTTC 45596
QY      90 ----- 90

```

Db 45595 TCTAATTGATGCTTTTTCACAGATTATACCTTTTGGGAATCAAGTCCAGGT 45536
 QY 90 ----- 90
 Db 45535 TCACAAAACCTTGANGANGCTGTGTTAGTCCAAAAATTTCTGTGTAATCTGTTTTTTT 45476
 QY 90 ----- 90
 Db 45475 TTCTCATGTACAAATCAAAAGTGAACCTAGTTTTTTCTATATATACGTGCTAGCTT 45416
 QY 90 ----- 90
 Db 45415 AAGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAAATTACTTGTTATATCGAGCT 45356
 QY 90 ----- 90
 Db 45355 CGCGCAAAAGCCAAACACAGAGAGCTCGCAAAATTTGATGTTAAAGCATATATATATC 45296
 QY 90 ----- 90
 Db 45295 TTAGCAATGAGCTCTGCAAAAAGATATACATCAAAACATGTTTACTCTGCTTTATGCGAA 45236
 QY 90 ----- 90
 Db 45235 GAAGTTTATCCGATTTGCTGTTATCTGCAAACTTGATCCGCGTTTAGAGTCTGATT 45176
 QY 90 ----- 90
 Db 45175 TCAGTTTCGCTCAGATGTTAAATCTCACAGCTTGAGTATGATGCTGCTACTCCA 45116
 QY 90 ----- 90
 Db 45115 CTCGAAATTCGCAATGCTGAAAATTTGGAAGCAGCAAACTTATCATGCGCCAAACAGA 45056
 QY 90 ----- 90
 Db 45055 TAAGAAATTTTGAGTTTAAAGTTCAGTTTCTCGCAAAATCAACCGGTGAGAAAT 44996
 QY 90 ----- 90
 Db 44995 TTGTCTATGCGGTACGTAGATATCAATTAATGATCTGCGAAACAGAAAATTTCTGGCAA 44936
 QY 90 ----- 90
 Db 44935 CCGTATCGTCTAATCTTAATATGCTGCGACATATCCATATCAATGAGGAGCTTTT 44876
 QY 90 ----- 90
 Db 44875 GGGCTTGTGTCGATGCTTACAGAAAGCAAAATTAAGTGTGTTAAGCGCGCTAAC 44816
 QY 90 ----- 90
 Db 44815 TCTACCAATCAGAAACGCTTTATTCGAGAACCATGTTGTTCTCAATTTCCATCCCTA 44756
 QY 90 ----- 90
 Db 44755 CGTACATATGGGCTTTCCCATTTGATGTTCTTAGAAGTTGACTTTCTCACAAACATTC 44696
 QY 90 ----- 90
 Db 44695 TGGGATTTGCTACAGTTGCAAGTGGAACAAGAAATATTCATTTGACGCGACACTCAATTT 44636
 QY 90 ----- 90
 Db 44635 CATGTCTTCCGCGCTTTGAGTACTTAAATGAGTGTTCAGGATGTTGTTTATTGGGAC 44576
 QY 90 ----- 90
 Db 44575 TCAGCAATGATAAAGACAGCAATGTTTTCAGAAAATGATTAATGCTTTCCGCGCTTGG 44516
 QY 90 ----- 90
 Db 44515 GGGACTACATACTTCTCTTAACAGCAAAATCTGATGTTTATGAGCAAGAAACTATTAACA 44456

QY 90 ----- 90
 Db 44455 CTGATTTAATAGAGAAAGAGAGATGTTTCTCTTCCGCGCAATTTTATTTGTTAGAA 44396
 QY 90 ----- 90
 Db 44395 AGCAATGATATGAATGCTGTGCTAGTGTAGTGAATTAATTAATGATTTGTGTGTTT 44336
 QY 90 ----- 90
 Db 44335 AGTTTCTTGATGTTTGAATGCTTTATTTCTTGCAACTATCTGGGATGATGCTCTT 44276
 QY 91 ----- 103
 Db 44275 ATGCACTTTTCTTGTTCGAGCAAAACCTTTGAGAGCTATAGGTGGGTGGGAAATCGGC 44216
 QY 104 ProAmValSerLyAspSerLeuGluGluGluPheGlyLysIleGluAsp 123
 Db 44215 CCTAATGCTCCAAAGATGACCTGAGAGAGAGTTCAGCAACTTTGGGAAAATCCAGGAT 44156
 QY 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAla 143
 Db 44155 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTCAATGATTTATATGAGATGATGATGCT 44096
 QY 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyLysSerPheLeuArgValAspPhe 163
 Db 44095 TTACAGGCTTAAAGACATGATGAAAGCCATAGGGGTGAGCTTTTGGGTGTGATTTT 44036
 QY 164 LeuArgSerGlnAlaProLysLys----- 171
 Db 44035 CTCGGGTCAAGCCGCCAAAAAAGTAAGACCTCTTGCGCATTTGATTTTAACTTTGA 43976
 QY 171 ----- 171
 Db 43975 AAACGCTCCAGTAAACATTTTGTAGTTTCATAATTTGCGTCAAACTAGTAGCGCTGAG 43916
 QY 171 ----- 171
 Db 43915 CTCTGTCTTGCCCTTACAGACAGATTAATCTGCTCTCTATTTCAATGTATGAGTAGGCT 43856
 QY 172 ----- 176
 Db 43855 CAACCTTTAAGTCTGAATCAAGTACCTTTGTGTATCTTCAGGAACAAATGGGCTGGC 43796
 QY 177 SerTyrAspAsnArgAsnGlyLysMetAsnHisLysProGln----- 190
 Db 43795 TCTTACGATTAACAGAAATGCAATATGAATCATTAACGCA -GGTTAGCTTGAATGTTG 43737
 QY 190 ----- 190
 Db 43736 AAGATGCTCTTGTACTAGATATGATATAGTACTGCTGTTTGAAGCTTTTGTATA 43677
 QY 191 ----- 207
 Db 43676 TTCTTACAGTATCTTCACTCAATATGAAGCTTTAAAGAGATGTCCAGCAAGTAAAGTT 43617
 QY 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
 Db 43616 CTGTGGAATGGGTTCCCTCTCTACTGCTCAACAAATCAATGAGACAAATTTCTGCACAT 43557
 QY 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
 Db 43556 GCGATGATATCTCTTGTGAGATTCAGAGGGGTAAAGTTAAAGTCCATCAAGAAATTTTGA 43497
 QY 248 LeuValIleuPheArgSerAlaGluGluAlaArgLysCysLysGluGlyLeuGlnGlyArg 267
 Db 43496 CTTGTGAGTTTAGAGCCCGAGAGAGCTCCCAATGCAAGAGAGGCTTACAGGGGAGG 43437
 QY 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
 Db 43436 TTATTCATTAATCTTAGAATCAAAATTAATGATCAAAAGATGAGTGTGCTCTGAGCA 43377

QY 288 AspAspThrSerPheTySerGlyMetLeuArgSerArgThrAspMetPheAsnAsp 307
DB 43376 GAGCATACGATGTTTACTCTGGATGAAACGGTCAAGGACAGATATGTTCAATATGAT 43317
QY 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
DB 43316 CCTCATTTTATCTCTCTCCATCTTCAATGAAATCTGGGCTATAGAGGCCCCCTCAGA 43257
QY 328 GlyThrAsnGlyLeuArgSerTyraAsnGlyValGluTyraAspValValGlyLeuGluPro 347
DB 43256 GGTACCAATAGAGGCTTCAATATATGAGGAAATACATGACCTGTGGTAAGAGGCCA 43197
QY 348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
DB 43196 AACTGAGAGAGGCCATCTGCMAATGGAACCTGCAATCTCCATCTCCACAGACCTGGA 43137
QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerArgProAspSerTrp 387
DB 43136 ATCTCCCATCTCTCTGACAAAGGTACGAGGCCCTATAGGTCATAGCCCAATCTTGG 43077
QY 388 GluGlyTyraAspProAlaGlnLeuValArgGluSerLeuArgThrArgArgAspGlySer 407
DB 43076 GAAGCATATGATCTCTGCTCACTGGTGCAGAGAAAGTAAACGAACCAAGAGATGATCA 43017
QY 408 ValAspGlyPheThrProMetGlyValAspGlyArgSerPheGlyArgGlySerValAla 427
DB 43016 GTGAGCGGTTTACTCCAAATGGGTGTCATAGAGAGGTCAATTTGGTGAAGTTCAAGTGT 42957
QY 428 AlaArgProIleArgGlyLeuProAspSerArgHisIleTrpArgGlyMetIleAlaIys 447
DB 42956 GCTAGACCTATCCGTGAGCCCTGATGTCATGATGACATATGAGAGAAATGATGGCAAG 42897
QY 448 GlyGlyThrProValCysCysAlaArgCysValProMetGlyLeuGlyIleGluThrIys 467
DB 42896 GGTGGAACCTCCGCTGTTGTGCTGCTGTGTATACCTATGGAAGGGATTTGAACATAA 42837
QY 468 Leu----- 468
DB 42836 CTGTAGATATATTTCTAGACATTAAACCTTCTAGTGTCTTTTCAAGCGAATT 42777
QY 468 ----- 468
DB 42776 ATATATTTCCATTTCATTCTGATGGAAGTAACTATTATATAGATGATGATTTTAT 42717
QY 468 ----- 468
DB 42716 TTACTATTAATTGTTTATGTTCTGAGATGCTTGATTTTCATGGTTCATTCTTTG 42657
QY 469 -----ProGluVal 471
DB 42656 GCATGGCCCTCAATTACTGACTTGTGTTTTTTTAAATTAATTAATTAAGGCTGAGGT 42597
QY 471 lValAsnCyseSerAlaArgThrAspLeuAsnMetLeuAlaIysHisIleValAlaIle 491
DB 42596 CTTCAATTTCTTCAAGCAAGACGATTTGATATATGCTCGCTTAACATACGCGGTGCAAT 42537
QY 491 eGlyCysGlnIleValPhePheValProAspArgGluLeuAspPheAlaSerTyThrGly 511
DB 42536 TGGAGTGTGAGATCGTTTTTTTTCGTACAGACAGGAAAGATTTTGGCTTCACTGCA 42477
QY 511 uPheLeuArgTyrlLeuSerSerLeuAspArgAlaGlyValAlaIysLeuAspArgGlyTh 531
DB 42476 ATTTCTCCGGTACCTTAAGCTCAAAAGATCGGGGGGTGTGGCAAAATTAATGATGTGAC 42417
QY 531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGly 551
DB 42416 AACTTATTTCTTGGTGCCTCCATCAGATTTCTTAATCTATGATCTCCAAAGTGAACCGGCA 42357
QY 551 nGluArgLeuTyrlGlyValValLeuIysLeuProProProAlaValProValThrAlaSe 571
DB 42356 AGAAGGCTATATGTGTGTCTCAAGTTACCCCGCCAGCGCGTCTCTGTTACAGCATC 42297
QY 571 rTyraArgGlnIysSerGlnSerAsnProLeuHisIleTyraMetArgGlnAlaArgAspSerPr 591

DB 42296 ATATGACAGAGATCTCAGTCCAAATCTCTGTGATTAATGATCAAGCCGGGATTCACC 42237
QY 591 oAlaAsnAlaSerHisSerLeuTyrlProProArgGluAsnTyrlIleArgGlyAlaProGly 611
DB 42236 TCCCATGTGTATGTCAGATTATATCTCTTAAGGGAATAATTAATTAAGGGGTGACCA 42177
QY 611 uHisLeuThrAlaAlaSerTyrlProSerValSerGluProLeuArgIleProAsnAsnAl 631
DB 42176 ACATTTGACAGCTGCTCAAAACCAATCTGTATGACGAGCTCTCAGAAATTAATTAATG 42117
QY 631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLe 651
DB 42116 AGCGCTTAAGCTGGGGTTAGTTTACTCCGAGCTTTTACCACTCTGGGATCTATCT 42057
QY 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh 671
DB 42056 CCTTCACACTTCTCAACTGCTGCGCCCTGAGAGTACACCAACTATATGTCAGGACCTTCAAC 41997
QY 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyraAsnGlyGluAlaProSerGlnAl 691
DB 41996 AGTTGTTCACAGACATCTCAGTCCAAATGACTGTATCAATGAGAGAGACCGTCTCAAGC 41937
QY 691 aTrpLeuArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrl 711
DB 41936 TTGGAAAGAGGTCCACAAACAGTTCAATGATGCTCAATATGATCATTCACAAATACGG 41877
QY 711 yAsnGlnTyrlThrProAlaGlyIleLeuProProProProSerArgTyrlProProAlaSe 731
DB 41876 AAATCAGTACACTCAGCTGAGGCAACTACCTCTCTCTGCGGTTACCTCCAGCTTC 41817
QY 731 rAsnAsnProAsnTyrlThrSerGlyMetValHisGlyAsnMetGlnTyrlGlnSerGlnSe 751
DB 41816 AAACAAACCCCACTACACTATGTAATGTGCAATGTGCAAGCAATGAAATACAGAGCCATC 41757
QY 751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrlSerMetTy 771
DB 41756 TGTTAACATGCTCAGCTGCTCCGTTACCAAAATATGCTCATATATATTCATGTA 41697
QY 771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrlGlnProGly 791
DB 41696 CACTCAGGGTTCGTCAATATCACTCTGTTCTCAGCCCATGCTCAGCAAAATACCAACAGA 41637
QY 791 uAlaSerMetProAsnGlnAsnTyrlGlyProIleProSerTyrlGlnGlnAlaAsnPheHis 811
DB 41636 AGCTTCATGCTCAACCAAAATCTATGCTCCAAATTCAGATTTATCAGAGCTTAATTTCA 41577
QY 811 eGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGly 831
DB 41576 TGGCGTAAACAATATCAGGACAGACCTTAACCTTCCCAATTTCAACCTCCATGCA 41517
QY 831 nProProAlaAspValAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
DB 41516 ACCCAACAGAGATTAAGCAAAATTTAGGCCAACAACCAAGCACTACGATTTGACGCTAT 41457
QY 851 tIleSerGlyAspGlyGlnGlyThrThrAspArgGlyValAspLeuAsnGlnAlaArgTyrl 871
DB 41456 GATCTCTGGGAGATGCTAGGGTACACAGATGGGAGGTGATTAAGATATAGATATCCA 41397
QY 871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGlnGlnIleArgGlnGlnGln 891
DB 41396 GTCAAACACTCAATTTGACAAACCTTCTTCTCCAGATACAGACAGAAACAGACAGCA 41337
QY 891 nSerSerGlyThrProAlaGlyGlnGlyPro 901
DB 41336 GTCTTCAGGTATCTCGGCTGGACAGGGGCT 41306

RESULT 7
AX420422/c 801 bp DNA linear PAT 18-JUN-2002
LOCUS AX420422
DEFINITION Sequence 4 from Patent W00212518.
ACCESSION AX420422
VERSION AX420422.1 GI:21524578

KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1
AUTHORS Amao, R.M., Schomburg, F.M., Michael, S.D. and Patton, D.
TITLES Floral induction gene
JOURNAL Patent: WO 0212518-A 4 14-FEB-2002;
WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES
source 1.801
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
/note="FPA antisense fragment"

ORIGIN
Alignment Scores:
Pred. No.: 1.35e-49 Length: 801
Score: 1436.00 Matches: 265
Percent Similarity: 99.62% Conservative: 0
Best Local Similarity: 99.62% Mismatches: 1
Query Match: 29.91% Indels: 0
DB: Gaps: 0

US-09-920-705-3 (1-901) x AX420422 (1-801)

QY 197 AAGPhlyaglyAaPValGlnProSerLysValLeuTrpIleGlyPheProProthAla 216
DB 799 GACTTAAAGAAATGTCAGCCAGTAAGATTCGTGGATTCGGTTCCTCTCTGCT 740
QY 217 ThGlnCySaSaSaSpGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
DB 739 ACACAATGCAATGATGAGCAAAATTCGCACAATGCCATGATCTCTTGGTGAGATCGAG 680
QY 237 ArgVallySerGlyProSerArgAsnPhaIleuValGluPheArgSerAlaGluGlu 256
DB 679 AGGGTAAAGATTAACCATCAGAGAAATTTTGCATCTGTGAGCTTGGAGCGCGAGGAA 620
QY 257 AlaArgGlnCySaLysGluGlnIleLeuGlnIleValArgLeuPheAsnAnProAqIleLysIle 276
DB 619 GCTCGCAATGCAAGGAGGCTTACGAGGAGGTATTCTCAATATCTTAGAATCAAAATT 560
QY 277 MetLysSerAsnAspGluLeuProProGluGlnAspAspThrSerPheLysSerGlyMet 296
DB 559 ATGTACTCAAAACATGATGTTGCTCTCTGAGCAAGACATATCTATCTCTGATG 500
QY 297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCyValSerSerProHisSer 316
DB 499 AAACGGTCAAGGACAGATATGTTCAATAAATGATCTTCAATTTCTCTCTCATTTCT 440
QY 317 ThGlnIleLeuProGlySerMetArgProLeuArgGlyThrAsnGluArgSerGlyAsnGly 336
DB 439 ACTGGAATTCCTGGGTCTATGAGGCCCTTCAAGAGTCAATATAGCGTTCAATAAATG 380
QY 337 AlaGluTyrAsnAspValIleGlyLysGluProAsnTrpArgArgProSerAlaAsnGly 356
DB 379 GCGAATACATATACGTTGTTGGTAAGAGACCAAACTCGAGGAGCGCATTCGCAATG 320
QY 357 ThGlnIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376
DB 319 ACTGGAATATCTCCATCTCCAAACGAGACCTGGAATCTCTCCATCTCTCGACAGGATG 260
QY 377 ArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuVal 396
DB 259 AGGGGCCCTAAGAGGTAAACCCCGATCTTTGGAGAGATATGATCTCTGCTGATGGTC 200
QY 397 ArgGluSerLysArgTrpArgArgAspGlySerValAspGlyPheThrProMetGlyVal 416
DB 199 AGAGAAAGTAAAGAACCAAGAGAGATGATGATGAGCGGTTTATCTCCAAATGGGATCT 140

QY 417 AAGPhlyaglyAaPValGlnProSerLysValLeuTrpIleGlyPheProProthAla 436
DB 139 GATGAGAGGTCATTTGGTTCAGAGTTCAAGTCTGCTGCTGAGCTTCTGCTGCTGAT 80
QY 437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyIleThrProValCysCysAlaArg 456
DB 79 TCGATCACATATGAGAGCAATGATTCGCAAGGAGGAACTCCGCTGTTGTCTCTCT 20
QY 457 CyValProMetGlyLys 462
DB 19 TGTGTACCTATGGAAG 2

RESULT 8
AK065222
LOCUS
DEFINITION
ORYZA SATIVA (JAPONICA CULTIVAR-GROUP) cDNA clone:J013002620, full insert sequence.
ACCESSION
AK065222
VERSION
AK065222.1 GI:32975240
KEYWORDS
FLI, cDNA, CAP, trapper.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1
The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-Length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shienhiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ohtomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurotaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Mikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Mikura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishi, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Oosato, Y., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
Science 301 (5631), 376-379 (2003)

JOURNAL
MEDLINE
22752273
12869764
PUBMED
REFERENCE
2 (bases 1 to 3543)
Adachi, J., Aizawa, K., Akiyama, T., Aizawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imanura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa, Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kuritani, C., Kurotaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Mikura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Mikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Oosato, N., Ota, Y., Ohtomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shienhiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahata, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.
Direct Submission
Submitted (05-DEC-2001) Shoenji Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of

COMMENT
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skluchi@ias.affrc.go.jp; Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28k full-length cDNA clones from japonica

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
Negata, T., Kawasashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shiehiki, T. and
Yamamoto, M.
FAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Mitra, J.,
Mizuno, K., Nariikawa, R., Nikiura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hangaki, T.,
Hara, A., Hahizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imocani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Kaneko, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,
Koyu, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Oseto, N.,
Oca, Y., Satoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shihagawa, A., Shiraki, Y.,
Sogaabe, Y., Tagami, M., Tagami, Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

```

FEATURES
    source
        location/Qualifiers
            1..3543
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="rRNA"
                /cultivar="Nipponbare"
                /db_xref="taxon:39947"
                /clone="J013002620"

```

ORIGIN

Alignment Scores:

Pred. No.:	4,598-46	Length:	3543
Score:	1365.00	Matches:	348
Percent Similarity:	52.55%	Conservative:	150
Best Local Similarity:	36.75%	Mismatches:	303
Query Match:	28.43%	Indels:	146
DB:	8	Gaps:	36

US-09-920-705-3. (1-901) x AK065222 (1-3543)

QY 16 GlnSerAsnAsnLeuTrpAlaGlySerLeuThrProGluThrGlnSerAspLeuThr 35
Db 245 GAGACGAACACCGCTGTGGTGGCACTGCCCGGCAGGCGCGGAGACGACTGTATG 305
QY 36 GluLeuPheGlyArgTyrcIyAspIleAmpArgIleThrValTyrSerSer--ArgGly 54
Db 305 GCGTCCTTCTCCCGCACGCGCGGCTGCATGCGTCAGTAGCGCCGCGCGCCCGCAGC 365
QY 55 PheAlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGln 74
Db 365 TACGGCTTGCTGTGCTGTCCGCTCCGTCGCGAGCGCCGCGCCCTCGACGCGCTCAG 422
QY 75 GlyAlaAsnLeuAmGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCys 94
Db 425 GGCTCCAAAGTAAGAGGACGCTCGTCCGGCTCGAGTTGCGCCGCGCGCTAGAGCTTT 484
QY 95 LysSerLeuTrpValGlyGlyIleGlyProAsnAlaSerLysAspAspLeuGluGluGlu 114
Db 485 AAGAATTATGAGTTGGTGGCATTTGCTCCCAATTATAAGACAGCACTCGAGAGAGAG 544
QY 115 PheSerLysPheGlyLysAlleuLysPheArgPheLeuArgIleuArgLysThrAlaPhe 134
Db 545 TTTAAGAAATTTGGAAAGTTGATGATGATTCCTTCGTATGACGACTTCGCGCTAC 604

135 ILeAspTYrTcgluMetkAspAlaIuGlnAla---LyseSerMetkAnglyLysPro 153
 605 ATTTGATTTTGACAAAGCTGAGAAATGCAATTTTCGGCACATAGAACCTTGAATGGAAAGATT 664
 154 MetGlyGlySerPheLeuAlaGlyValaAspPheLeuAlaGlySerGlnAlaProLysLysGlnGln 173
 665 TTAGGTGGACAAGAAACTGTGTGTGATTTCCAAAGGTCCAGGGGC-----AGAGCGGAA 718
 174 TTP-----AlaGlySerTYrAspAsnAlaGAsnGlyAsnMetAsnHisLysProGlnTYr 191
 719 TGGTTGGAGACAGGACGCTTCATGTGTAACTGACCTGGCAAG----- 763
 192 ProHisSerTYrGluAspPheLysGluValaGlnProSerLysValLeuTrpIleGly 211
 764 -----GGTTATGGAGTACGAATGAGAAATGCAATCCAAATGCTTTGGGTAGAGT 817
 212 PheProProTrpAlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeu 231
 818 TTCCCTTAATACAGCTTAA-----ATCAATGAGGAGGCACCTTCGGCAACGCAATGGCTGTGA 871
 232 PheGlyGluIleGluAlaGlyValaLysSerTYrProSerArgAsnPheAlaLeuValaGluPhe 251
 872 CATGGTGCCTTTACAAATACAAAGATTTCCTCCAAACAGGCAATATGCTTTGTGGAGTTT 931
 252 ArgSerAlaGluGlnAlaArgGlnCysLysGluGlyLeuGlnGlyArgPheAsnAsn 271
 932 GCACACAGTTGGGGAAGCTCTTAATGTAAGAAATACTGATGTCGTCTTTTCATGAT 991
 272 ProAlaGlyIleLysIleMetTYrSerAsnAspGluLeuProProGluGln---AspAspThr 290
 992 CAGAGATTTCAATATCTTTCTTCGAAACAGTAGCTTGGCCCAATTAATTAATATATCA 1055
 291 SerPheTYrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAspProSerCys 310
 1052 ACAGACAGTTTCTGGATTCCAAATCA-----GAGATGTATTAATGATGATGTCAGATAT 1100
 311 ValSerSerProHisSerThrGlyIleProGlySerMetAspProLeuAlaGlyThrAsn 330
 1106 GGGCGTTTCTGATTAT-----TTTGATCTTCGACGGGGGAAGATCA 1144
 331 GluArgSerTYrAsnGlyAlaGluTYrAsnAspValAlaGlyLysGluProAsnTrpArg 350
 1145 AGATAC-----TTTGAGTACAGTGTGTC----- 1166
 351 ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPro 370
 1169 -----CCTGTTTCTGGTGGTATTTCTTCCA 1192
 371 SerProAlaGlnGlyThrArgAspPheMetLysSerAsnProAspSerTrpGluGlyTYr 390
 1193 TCACCGGAATCTGGA-----AATCTCTCTTAATCTGACGATCTGCTCAAGATACATTC 1246
 391 AspProAlaGlnLeuValArgLysSerLysArgThrArgArgAspGlySerValAsp--- 409
 1247 GATCCA-----AGAGAAGCCAAAGAGTTGGGTTGGANGCAGACAGTCAACCT 1299
 410 -----GlyPheThrProMetGlyValaAspGluArgSerPhe 421
 1295 TATGATACAAAGGCGAGTAGTGAAGGCTTTATTTCTGTGGGTACTTCAACCGCGAGAGT 1354
 422 GlyArgGlySerValAlaAlaArgPro-----IleArgGlyProProAspSer--- 437
 1355 GCAGCTTGTGAGAGAGAGCTCAAGTCTGCTATTGGATTCATGGCACAGTGCACCGAACA 1414
 438 -----AspHisIleTrpArgGlyMetIleAlaLysGlyIleProValCysCys 454
 1415 TCATACCTTTGAGCACTTCTGGCGTGTAGACATTGCCAAAGGTGATCTCTGTATGTGCT 1474
 455 AlaArgCysValaProMetGlyLysGlyIleGluThrLysLeuProGluValaValaAsnCys 474
 1475 GCTGGCTGTGTGCTTAATGAGAGAGGTGTAGATACCTTTAACCGAGATGTTGTAAATGTT 1534

QY 475 SerAlaArgThrAspLeuSerMetLeuAlaLeuValHisTyrAlaValAlaIleGlyCysGlu 494
 DB 1535 TCGGTAAGTAAGTGAAGTATGCTGGCAAGCACTTATAGAGACGCTTCAGGGTTGGAT 1594
 QY 495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
 DB 1595 ATGCTCTTCTTCTTGGCAGATAGTAAGATGACTTTGTTCTTACACTGAATTTTGGCG 1654
 QY 515 TyrLeuSerSerTyrAspArgAlaGlyValAlaLeuLeuAspAspGlyTyrThrLeuPhe 534
 DB 1655 TACTTGGCTCAAAAGTGGGCGAGGGTGTAAAGTTGATGGGGGACTCTCTTATATTT 1714
 QY 535 LeuValProProSerAspPheLeuThrAspValLeuGluValThrArgGluGluArgLeu 554
 DB 1715 TTGGTCCACCTTCCGATTTTTCAGAAATGTTTTCAGACTCATGTGTCAGAAACGCTTT 1774
 QY 555 TyrGlyValValLeuAspLeuPro-----ProProAlaValProVal 568
 DB 1775 TACGGTGTAGTATTACACATTCGCAAAATGTCGTGCTGCTGCTGCTGCTGCTGCTGCT 1834
 QY 569 ThrAlaSerTyrArgGluGlu-----SerGlnSerAspProLeuHisTyrMetAsp 585
 DB 1835 CCTGCTGTACAAAGGCGCACAACTACTGACACCAAAATCACAACT-----TACTACGAT 1888
 QY 586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyr 605
 DB 1889 GAA--AGGGAATTTCTTTGCGAGAGAGGTATATGATATGATTCGCTTACCAACATCAC 1945
 QY 606 IleArgGlyAlaProGluHisLeuThrAla----- 615
 DB 1946 CATCGAGATGCT--GATCACCGTGGATCTTTACGTGAGGATTCATTGCATCAATGGGG 2002
 QY 616 -----AlaSerTyrProSerValSerGluPro-----LeuArg 626
 DB 2003 CAATTTCTTGCACGACCTCGGGGATGAGAGCAAGTAGTTCAGCCAACTTGGCCGGT 2062
 QY 627 IleProAsnAsnAlaIleProGlnAlaGlyValSerLeuThrProGluLeuAlaIleThr 646
 DB 2063 ATCCCTTACAAACAGAGCACTACAGGTACAGCCCTTCAGCTGATATGATAGCCACT 2122
 QY 647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMet 666
 DB 2123 TTATGTAACCTTTTGGCAGATGGGAGTATCATGCTGTGCTGCTGCTGCTGCTGCTGCT 2182
 QY 667 SerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGlu 686
 DB 2183 AGC-----TCAACAGACCAACCTGCATTAAACCAATGATGAT-----GATGATGCT 2227
 QY 687 AlaProSerGlnAlaThrPheAspArgGlyProGlnThrValHisAspAlaSerAsnGlnSer 706
 DB 2228 ACTCTTGCAGAAATATAGAGAA-----CTGAAATATCAAGCAATGCTTCTTACTTCATCC 2281
 QY 707 PheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProSerArg 726
 DB 2282 TTATAGACAGATGCGTAAT--TTCAGACATTCAGACAA-----CAGTTCACACAA 2329
 QY 727 TyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGln 746
 DB 2330 CAACTGAGACAGTATCATGCAAACTAT-----GGGCACTTGGCAGAGTGCACAGAG 2383
 QY 747 TyrGlnSerGln-----SerValAsnMetProGlnLeuSer 758
 DB 2384 CACCCAAACAAATAGTGTCTTAAATCTGAAATGATGCTTAACTTGGCCACACACCA 2443
 QY 759 ProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHis 778
 DB 2444 CCACTCTTACAGGTGCGCACCACTTCTGCTATATTCATCGCAAGTGTGTCACACTTGG 2503
 QY 779 ProValSerGlnProMetValGlnGlnTyrGlnProGluAla--SerMetProAsnGln 797
 DB 2504 CCT--ACGCAAGATGAGTCAAGCAAGTATCAACAGAGCAGATATATATGACTCAAGC 2560
 QY 798 AsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGly----- 812

DB 2561 AATTATGTCATTGTGGCCACACTTAGCACTTCTTATCTTACAGCGCACATCAACAATT 2620
 QY 813 ValTyrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaIleMetGlnPro 832
 DB 2621 GTTGGACATCCGCTGACACAGCTCCTGTTGCTGGCAGATTCCTCTCTATGCAAGCT 2680
 QY 833 ProAlaAspValAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIle 852
 DB 2681 CCTGCT-----GCCGCAACACTCTCTGT 2704
 QY 853 SerGlyAspGlyGlnGlyTyrThrAspGlyGluValAspLeuAsnGlnArgTyrGlnSer 872
 DB 2705 GCT-----GCACAGCTTCTGCTGAC--GAGCGCAAGACGAAACAGAAATGACGGG 2755
 QY 873 ThrLeuGlnPheAlaIleAsnLeuLeuGlnIleGlnGlnLeuGlnGlnGlnSer 892
 DB 2756 ACTCTCAGTGGCTGACGCGCTGTTGGTGCAGTACACAGAAAGCTGGAATCAACT 2815
 QY 893 SerGlyThrProAlaGlyGln 899
 DB 2816 TAAGCTATCTGGAATCA 2836
 RESULT 9
 AP005735/c 124083 bp DNA linear JUN-28-2004
 LOCUS AP005735
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
 BAC clone:OSUNB0034B12.
 ACCESSION AP005735
 VERSION AP005735.3 GI:50726236
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 BUKARYOTA; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE 1
 AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, BAC
 clone:OSUNB0034B12
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 124083)
 AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
 TITLE Direct Submision
 JOURNAL Submitted (11-Sep-2002) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@ias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 On Jul 27, 2004 this sequence version replaced gi:40736955.
 Genes were predicted from the integrated results of the following:
 GENSCAN (http://CCR-081.mil.edu/GENSCAN.html), FGENESH
 (http://www.softberry.com/), GeneMark.hmm
 (http://opal.biology.gatech.edu/GeneMark/), GlimmerX
 (http://www.tigr.org/tdb/glimmerx/glmr_form.html), RiceHMM
 (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
 (http://bioinformatics.iastate.edu/cgi-bin/bp.cgi), slm4
 (http://globlin.cse.psu.edu/html/docs/slm4.html), gap2
 (http://www.tigr.org/software/glimmer/), BLASTX and BLASTN.
 The genomic sequence was searched against NCBI Nonredundant Protein
 database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
 sequence database at RGP or DDBJ. Protein homologies of the coding
 regions were searched against NCBI Nonredundant Protein database
 with BLASTP. ESTs represent the identified cDNA sequences using
 BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
 Full-length cDNAs represent the identified cDNA sequences using
 BLASTN with the corresponding DDBJ accession no.
 A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without
 significant homology to any protein but with full-length cDNA or
 EST homology (covering almost the entire length of partial
 sequence) is classified as an 'unknown' protein. A gene predicted

by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OSJNB0034B12 has an overlap with P0450E05 (DDBJ: AP006756) clone at 5' end and with P0217C03 (DDBJ: AP006173) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://ftp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES
Source

1..124083

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="Genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="9"

/clone="OSJNB0034B12"

3185..6071

/gene="OSJNB0034B12.1"

join(3185..3279,3941..4032,4234..4283,4368..4450,

4600..4708,5392..5493,5640..6071)

/gene="OSJNB0034B12.1"

/note="supported by full-length cDNA(s): AK103734"

join(3241..3279,3941..4032,4234..4283,4368..4450,

4600..4708,5392..5493,5640..5686)

/gene="OSJNB0034B12.1"

/note="contains full-length cDNA(s): AK103734"

/codon_start=1

/product="unknown protein"

/protein_id="BAD33813.1"

/db_xref="GI:50726237"

/translation="MAMMGLAKNPKMINHRYASDMOLHRLSPACVMEKOCYSR

ETKASVLAAGVAPICETETESTSLVSPDFVSVCTEDGLIKRYADPTTSIF

EYKESKVAAPLPGFRMKDPTKDCALNMAIGERRNOHETQPNCKICPGG

PRCIKXSEGTDL"

7566..12011

/gene="OSJNB0034B12.2"

join(7566..7812,7988..8078,8413..8491,8850..9183,

9374..10408,10581..10730,11221..12011)

/gene="OSJNB0034B12.2"

/note="supported by full-length cDNA(s): AK065778"

join(7786..7812,7988..8078,8413..8491,8850..9183,

9374..10408,10581..10730,11221..11655)

/gene="OSJNB0034B12.2"

/note="contains EST(s): C98201(C1066), D15674(C1066A)

contains full-length cDNA(s): AK065778"

/codon_start=1

/product="putative translation initiation factor IF-2"

/protein_id="BAD33814.1"

/db_xref="GI:50726238"

/translation="MAMRLRRKDPFHTGLVNLASHVDHGAGNFSSGTEGLADPVLS

DHTPIVKAANCYAKICTINFHAGVYMLMSKREAVGLAKPKKRVKRETR

QPEVAPYVAPROKIAIKSPKPTVDIFDGMWLLDLKRTGASGALGDTITDGEKY

EEFPDLSIDLAELVAMELGVNIRRMHGEGLTERPRAVVTIMGVHVDGLDLSLR

QNSVAKREGGITHIGAFVYMPESGATITFDTHGAFASMRARGAATDIDVLVY

AADDGMPTLEASHAKANPVIIVAVNCKDSQADPERVRIQSGSGALLIDMGD

VYVETISATYKLGIDLEBALLQAEIMDLKARIDGPAQAVPEARVDRGRPLATY

VASGLVSGOHTVGAEGWIRISLRDTAGKITESAKPAPVEIGELRLPAGDGVVY

VDSEBRLNSQGRKKQEDRLKIDEDMTEBEAIGETPERVEMTITIVADOGCV

QAVTDLRSINSPOVFVNIVHVGVPISOHNDIAOACRAIVTGNRTTPSAILTAA

TDANKILLHKVYHLLLEMGREIYERKAPGPPEYVSGEAEVNIIFELGSKSKGPD

IKIACRITDGLSTKTMRLIRSGDVFEPCASLKKRKQDAETVDKMGDCGLVID

CNDPQVGDIVQLEQVIRKPKFISTOSGAVREIC"

complement(12465..12713)

/gene="OSJNB0034B12.3"

complement(12465..12713)

/gene="OSJNB0034B12.3"

/note="start and end point are not identified"

complement(12465..12713)

/gene="OSJNB0034B12.3"

/note="predicted by FGENESH etc."

gene
mRNA
CDS

/codon_start=1
/product="hypothetical protein"
/protein_id="BAD33815.1"
/db_xref="GI:50726239"
/translation="NAGSKSVVILLAVLAVFAVLGAAAAAPAPAPASGGEAA
AAVAVNPAVAVEARRETVEMLMRLPAGSPKPGH"
13577..15039
/gene="OSJNB0034B12.4"
join(<13577..13620,13698..13844,14101..14263,
14944..15039)
/gene="OSJNB0034B12.4"
/note="start and end point are not identified"
join(13577..13620,13698..13844,14101..14263,14944..15039)
/gene="OSJNB0034B12.4"
/note="predicted by GENSCAN etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD33816.1"
/db_xref="GI:50726240"
/translation="MMKNLIPMPFVSNVQDPRIEFTISNRQAKERSYVVKY
PRMVEIGENRNNHESSELSJDLALIOREKDKTDIDLVQVODEIGWREDSKY
RLIAYEMIMHKKHMLVLAEYVKTSSNDGSGVQVQDEIRATTR"
complement(15987..19946)
/gene="OSJNB0034B12.5"
complement(join(15987..16346,16512..16613,16686..16771,
17095..17366,17870..17961,18055..18192,18269..18339,
18646..18829,18907..19039,19421..19936))
/gene="OSJNB0034B12.5"
complement(join(16003..16346,16512..16613,16686..16771,
17095..17366,17870..17961,18055..18192,18269..18339,
18646..18829,18907..19039,19421..19936))
/gene="OSJNB0034B12.5"
/note="supported by full-length cDNA(s): AK070794"
complement(join(16185..16346,16512..16613,16686..16771,
17095..17366,17870..17961,18055..18192,18269..18339,
18646..18829,18907..19039,19421..19515))
/gene="OSJNB0034B12.5"
/note="contains EST(s): AU085809(C53716), C28019(C53716)
contains full-length cDNA(s): AK066076, AK070794"
/codon_start=1
/product="putative tbc1 domain family protein"
/protein_id="BAD33817.1"
/db_xref="GI:50726241"
/translation="MSGGGGGGGGGGGGGGPNNTFRFNOTLRNVQMLKGRFPKRVL
LTRSEPLSPYSPRENDREYEQNESQEGKQAGNTADSMGAKSNPSTST
NSLPDAGLVSGARADDSARIAKFTNELSRPAVILDKRELGSQVPPMRPIWRL
LGYAPNADRRREGVLTFRKLEVECVSYDIDPTSRDEEIMRLROIADCPRTVPD
VTFPHQIOKSLERILTYWALRHPASGVVGGINDLPTPLVAFLESEHLEGMMDTSM
EKSPDQSNIEADCYWCLSKPLDGMODHTYPAQDIOQLVRLKELVHRIDEPYKX
MEQGLDFOFAPRMFCIMIRIPIPHVITRLMDYTLAAGDYLPDLVITISFLLTW
SDLKTLDPQEMVFIQHLPTFRWAHELEWLSRAYVMHTFKSSPSHLAS"
20737..21569
/gene="OSJNB0034B12.6"
join(<20737..20769,21306..21569)
/gene="OSJNB0034B12.6"
/note="start and end point are not identified"
join(20737..20769,21306..21569)
/gene="OSJNB0034B12.6"
/note="predicted by FGENESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD33818.1"
/db_xref="GI:50726242"
/translation="MSVATNLGHTHPAPSNFLIAGASRRATWENAEBCGGVSP
GAAGDEGSGDGDGVNAVYLFEPPTLTGSGDGVGEVDGATLLMAGLI"
complement(22328..22807)
/gene="OSJNB0034B12.7"
complement(22328..22807)
/gene="OSJNB0034B12.7"
/note="start and end point are not identified"
complement(22328..22807)
/gene="OSJNB0034B12.7"

CDS
mRNA
gene

gene
mRNA
CDS


```

Db      40877 ATCGAGAGTCT---GATCAGCGTGAGTCTTTACGTGAGATTCATTCATCAATTTGGGCGC 40821
Qy      616 -----AlaserlyProserValserGluPro-----LeuArgI 627
Db      40820 AAATCTTTGACGACCTCGGTGTGATGAGCAAGTAGTTCAGCCAAACCTTGGCCGTA 40761
Qy      627 leProAenAenAlaIAProGlnAlaIValserLeuThProGlnLeuLeuAlaThL 647
Db      40760 TCCCTACAAAGCAGACTACAGCTACAGCTTACAGCTTCCAGCTGATATATATAGCCACTT 40701
Qy      647 euAlaserIleLeuProAlaThserGlnProAlaIAProGlnSerHleGlnPromets 667
Db      40700 TAGCTAAACTTTTCCAGATGGGAGCATCATCAGCTCGGCACCTGGCCACTTCCCTGA 40641
Qy      667 ergIyProserThValValserThValserHleGlnSerHleGlnValserGlyGlu 687
Db      40640 GC-----TCAGAGACCAACCTGCATTAACCCAAATGAAT-----GATGCACTTA 40596
Qy      687 laProserGlnAlaThValserGlyProGlnThValHleAenAlaserAenGlnSerP 707
Db      40595 CTCCTGCAAAAGTATGAGA-----CTGAAATACAGCAATAGCTTCTTACTTCAATCTC 40542
Qy      707 heGlnGlnIyGlyAenGlnIyThProAlaGlyGlnLeuProProProProserArgT 727
Db      40541 TAGAGCAGATCGTAAT---TTCAGCATTCAGGACAA-----CAGTTACGACAAAC 40494
Qy      727 yrProAlaserAenAenProAenIyThserGlyMetValHleGlyAenMetGlnT 747
Db      40493 AACCTGAGAGCATTCATCTGCCAAATAT-----GGAACTTGGAGGAGTGACAGAGAC 40440
Qy      747 yrGlnSerGln-----SerValAenMetProGlnLeuSerP 759
Db      40439 ACCCAACACAAATAGTGTTCATCATCTGAAATGATCTAAATCTGCCACACACACAC 40380
Qy      759 roLeuProAenMetProHleAenAenIySerMetIyThGlnIySerSerAenHleP 779
Db      40379 CACCTCTTACGCTGCCACACAGTTCTGTATATATATATATATATATATATATATATAT 40320
Qy      779 roValserGlnProMetValGlnIyGlnIyProGlnIyA-----SerMetProAenGlnA 798
Db      40319 CT---ACGAGAGATGAGTACGACAACTATACACAGAGAGAGATTTATATAGTCAAGCA 40263
Qy      798 snIyGlyProIleProserIyGlnIyGlnAlaAen-PheHleGlyVal----- 813
Db      40262 ATATATGTCATTTGGCCAGATGAGTTCATATCTTCAAGTTCAGCAATTCATCAATCC 40203
Qy      814 -----ThThAenGlnAlaGlnAenLeuAen 822
Db      40202 CATTTCTTCTTATTTCCCAAGTAACTTGCACAAATATATCAATGAGGTTTCTTCT 40146
Qy      823 ProserGlnPheGlnAlaIaMetGlnProAlaIaSerPLeuAen----- 838
Db      40145 ---GCAGTGTTCAGAGCTCTTGACTTGGCTTATGATAGGTTAACTATAGGTTTCTCT 40089
Qy      839 LeuGlnIyProGlnAenGlnAlaIaLeuAenGlnIyProMetIleSerGlyAen----- 855
Db      40088 TTACAGGCGCACATCAACAAATGTGTGCACTCTGTGCACAGATCTCTGTCTGCG 40029
Qy      856 -----GlyGlnIy 858
Db      40028 CAATTCCTCTCTGATGAGACTCTGTGCGGACAAAGCTCTGTGCTGCACAAAGCT 39969
Qy      859 ThrThAenGlyGlyValaSerIyAenGlnIyArgIyGlnIyThLeuGlnPheAlaIa 878
Db      39968 TCTGTGTGAC---GAGGCAAGACGAAACAGAAAGTATCAGGAGACTCTCAATGGCTCAG 39912
Qy      879 AenLeuLeuGlnIleGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIy 898
Db      39911 CGCTGTTGGTGCACCTACACAGAGCTGGAATCAACTTAAGGCTATCTCGGAAT 39852
Qy      899 Gln 899

```

```

Db      39851 CAA 39849
RESULT 10
AC108755/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) chromosome 9 BAC clone
OSUNBa0028C22, complete sequence.
AC108755
VERSION
AC108755.2 GI:50345123
SOURCE
ORGANISM
Oriza sativa (japonica cultivar-group)
Oriza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoae; Oryza.
1 (bases 1 to 154228)
Lee,J.-S., Hahn,J.-H., Yoon,U.-H., Lee,K.-S., Kim,Y.-H., Eun,M.Y.,
Kim,H.-I., McCombie,W.R., Zutavern,T., de la Bastide,M.,
Spiegel,L., Muller,S., Nascimento,L., Balija,V., Bell,M.,
Miller,B., Katzenberger,F., Andrade,M.V., Dike,S., O'Shaughnessy,A.
and Palmer,L.
Oriza sativa (japonica cultivar-group) chromosome 9 BAC clone
OSUNBa0028C22, complete sequence
2 (bases 1 to 154228)
Hahn,J.-H. and Kim,H.-I.
Direct Submission
Submitted (31-JAN-2002) Rice Genome Sequencing Project, National
Institute of Agricultural Science and Technology(NIAST), RDA, 249
Seodun-dong, Suwon 441-707, Korea (E-mail:jhhahn@rda.go.kr,
Tel:82-31-290-0309, Fax:82-31-290-0308)
3 (bases 1 to 154228)
Hahn,J.-H. and Kim,H.-I.
Direct Submission
Submitted (16-JUL-2004) Rice Genome Sequencing Project, National
Institute of Agricultural Biotechnology (NIAB), RDA, 225
Seodun-dong, Suwon 441-707, Korea
On Jul 16, 2004 this sequence version replaced gi:18449960.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. OSUNBa0028C22
overlaps clone OSUNBa0065A15 (AC137594) from base 1 to base 12997.
The overlap is from base 155808 to base 168804 on OSUNBa0065A15.
OSUNBa0028C22 also overlaps clone OSUNB0034B12 (AF005735) from
base 102828 to base 154228. The overlap is from base 1 to base
51401 on OSUNB0034B12. This BAC clone was sequenced to phase II by
the Korean National Institute of Agricultural Biotechnology and was
completed to phase III by the Cold Spring Harbor Laboratory Genome
Center.
FEATURES
source
location/Qualifiers
1..154228
/organism="Oriza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/DB_xref="taxon:39947"
/chromosome="9"
/clone="OSUNBa0028C22"
/clone_11b="HindIII"
56468..56601
/note="The assembly is covered by high quality sequences
derived from a transposed plasmid subclone."
ORIGIN
Alignment Scores:
Pred. No.: 2,06e-34 Length: 154228
Score: 1117.50 Matches: 325
Percent Similarity: 45.76% Conservative: 114
Best Local Similarity: 32.40% Mismatches: 280
Query Match: 23.28% Indels: 264

```



```

Db      143527 TAGTAAACTTTGGCAAGTGGGAGCATGACGCTGGTCACTGGCCAGCTTCCCTGA 143468
Oy      667  englyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrrAsnGlyGluA 687
Db      143467 GC-----TCAACGAGACCAACCTGCATTAAACCAATGAAT-----GATGCATCTA 143423
Oy      687  lAProSerGlnAlaTTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerP 707
Db      143422 CTCCTGGCAAAAGTATGGAGA-----CCTGAATAATCAAGCAATGGCTTTACTTCACTCT 143369
Oy      707  heGlnGlnTyrrGlyAsnGlnTyrrProAlaGlyGlnLeuProProProProSerArgT 727
Db      143368 TAGAGCAATGGGTAAT-----TTCCAGCATTCAGAGCAAA-----CAGTTCAGCAAA 143321
Oy      727  yrProProAlaSerAsnAsnProAsnTyrrHisSerGlyMetValHisGlyAsnMetGlnT 747
Db      143320 AAGCTGAGAGCGTTCACTCTGCCAAACTAT-----GGAACTTGGCAGGGTGCACAGAGGC 143267
Oy      747  yrGlnSerGln-----SerValAsnMetProGlnLeuSerP 759
Db      143266 ACCCAACAACAACATAGTGTCTTACAATCCTGAATAAGACTTAATCTGCCACACACACAC 143207
Oy      759  roLeuProAsnMetProHisAsnAsnTyrrSerMetTyrrThrGlnGlySerSerAsnHisP 779
Db      143206 CACCTCTACGCTGCACCCAGTCTGTCTATATTATCATGCAAGTTGACACAGCTTTCG 143147
Oy      779  roValSerGlnProMetValGlnGlnTyrrGlnProGlnAla-----SerMetProAsnGlnA 798
Db      143146 CT---AGCAGATAGTACAGACCAACAGTATACACCAAGCAGTATTATATGACTCAAAAGCA 143090
Oy      798  enTyrrGlyProLleProSerTyrrGlnGlnAlaAsn-PheHisGlyVal----- 813
Db      143089 ATTATGTCTCAATTGGCCAGCATTTAGCAGTTCTTAATCTTCAGTCAAGCAATTCAACATCC 143030
Oy      814  -----ThrThraAsnGlnAlaGlnAsnLeuAsn 822
Db      143029 CATTCTCTCTTATCCCAAGTAAACCTTCCCAACAATATATAGTAGGGTTTGTG--- 142973
Oy      823  proSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysValAsn----- 838
Db      142972 ---GAGTGTTCAGCATGCTTGTGATCTTGCCTTAATGATAGGTAACATATGGTTTCTCT 142916
Oy      839  LeuGlnProGlnAsnGlnAlaAlaLeuAlaGlyLeuGlnProMetCileSerGlyAsp----- 855
Db      142915 TTACAGGCGCACCATCAACAATTTGTCGACCTCTGTCACACAGCTCTGTGCTGGCG 142856
Oy      856  ---GlyGlnGly 858
Db      142855 CAATCCCTCTCTGTATGCAAGCTCCTGCTGCCGCGCACAGCTCCTGTCCTGCACACAGCT 142796
Oy      859  ThrThraAspGlyGluValAspLysAsnGlnArgTyrrGlnSerThrLeuGlnPheAlaAla 878
Db      142795 TCTGCTGAC---GAGCGAGAACGAGAACAGAAAGTATCAGGACCTTCAAGTTGCTCG 142739
Oy      879  AsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 898
Db      142738 CGCTCTGTGGTCACTACACAGAAAGCTGGAATCAACTTAAGGCTATCTGGAAT 142679
Oy      899  Gln 899
Db      142678 CAA 142676

```

```

RESULT 11
AP006756/C 176553 bp DNA linear PLN 28-JUL-2004
LOCUS      Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
DEFINITION PAC clone: P0450E05.
ACCESSION AP006756
VERSION    AP006756.2 GI:50725340
KEYWORDS   Oryza sativa (japonica cultivar-group)
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

COMMENT

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

1
Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC clone: P0450E05
Published Only in Database (2004)
2 (bases 1 to 176553)
Sasaki, T., Matsumoto, T. and Katayose, Y.
Direct Submission
Submitted (31-MAR-2004) Takuji Sasaki, National Institute of Agricultural Sciences, Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://exp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jul 27, 2004 this sequence version replaced gi:46093787.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/geneMark/), Glimmer (http://www.tigr.org/tcb/glimmer/glimmer_form.html), RiceHMM (http://exp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/bp.cgi), sim4 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologues of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from 77 to 596 of the PAC clone. This sequence of P0450E05 has an overlap with OSUNB0055A15 (DDBJ: AC137594) clone at 5' end and with OSUNB0034812 (DDBJ: AP005735) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

gene
mRNA
CDS

```

1..176553
  /organism="Oryza sativa (japonica cultivar-group)"
  /mol_type="genomic DNA"
  /cultivar="Nipponbare"
  /db_xref="taxon:39947"
  /chromosome="9"
  /clone="P0450E05"
  1479..4246
    /gene="P0450E05.1"
    join(<1479..1885,3329..3366,3854..3945,4100..>4246)
    /note="start and end point are not identified"
    join(1479..1885,3329..3366,3854..3945,4100..4246)
    /gene="P0450E05.1"
    /note="similar to Arabidopsis thaliana chromosome 5, At5g47870"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="BAD34413.1"
  /db_xref="GI:50725341"
  /translation="MEALRSGPAAPPEYPAARVAAMGARRRGRSSSSVVAAMEG
GGGKGVPTTNYVPLDKATGMTPLVEIIRDLNRKVDKIIDPTNTVPWSPVLP

```


[illegible]

Db	162114	GCATTGCCAAAGGTGGATGTCCTCGTATATGTCGTCGCTGCTTGGCTATTAAGCAAGGGTG	1626535
Qy	464	leglunhrlylsleu-----	468
Db	162654	TTGAGATACCTCTTGTGATGTATACACTACACCTATCTTATCTGTCTCTACAGTTGATTC	1626595
Qy	469	-----	-ProgluValAlaAspCys 475
Db	162594	TATAGCCTCGATCTTACACATATTCACCTGTTTGGACCGGATGTTGTAATTTGTT	1626535
Qy	475	erAlaargThrAspleuAsenMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu	495
Db	162534	CCGCTGAACCTGGACATGATGATGCTGGCAACCATATATGAGACCGCTCAGGGTTTGATA	162475
Qy	495	leValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgT	515
Db	162474	TGCTCTTCTTCTTCCAGATAGTAGAAGAGACTTGTCTTACACTGAATTTTGGCGCT	162415
Qy	515	yrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheL	535
Db	162414	ACTTGAAGCTCAAAAAGTGGGAGGGGCTGTGAAGGTGATGGGGGACTACTTATTTT	162355
Qy	535	eulValProProSerSerPheLeuThrAspValLeuGlnValThrArgGlnIuAArgLeuT	555
Db	162354	TGGTCCACACTTCCGATTTTTCAGAAATGTTTTCAGATGTGATGTCAGAAACCTCTTT	162295
Qy	555	yrGlyValValLeuLysLeuPro-----	ProProAlaValProValT 569
Db	162294	ACGGTGTAGTATTTACACATTCGCGAAATGTCGTCGCTGCTCTCTCTCTGCTCCACTC	162235
Qy	569	hAlaSerTyrArgGlnIu-----	SerGlnSerAsnProLeuHisTyrMetAspG 586
Db	162234	CTGCGTACMAAGGCGACAACTPACCTGCACCGACGAAATCAACCT-----	TACTCGATG 162181
Qy	586	InAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgIuAsnTyrI	606
Db	162180	AA--AGGGAATTCCTTTCGACGAGAAGATATATGATTAAGCTTAGCAACAATACCC	162124
Qy	606	leArgGlyAlaProGluHisLeuThrAla-----	615
Db	162123	ATCGAGATGCT--GATCACCGTGATCTTAACTGAGATTCATTCATCAATTGGGGC	162067
Qy	616	-----AlaSerTyrProSerValSerGluPro-----	LeuArgI 627
Db	162066	AAATTCCTGCACGACCTCGGGTGATGAGGACAAATGATTCGACGCCAACCTTGGCGGTA	162007
Qy	627	leProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrL	647
Db	162006	TCCCTACAAACGACGAGCTACAGGTACGACCTTCACCTCACCTGATATGATAGCACTT	161947
Qy	647	eulAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetS	667
Db	161946	TAGCTAAACTTTTGCCAAAGTGGGCGAGTCATGACCTCTGGTACCTGGCCAGCTTCCCCTGA	161887
Qy	667	erGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyIuA	687
Db	161886	GC-----TCAACGACCAACCTCGATTAACCCAAATGAAT-----	GATCATCTA 161842
Qy	687	laProSerGlnAlaTyrLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerP	707
Db	161841	CTCTTGCAAAAGTATGAGA-----	CCTGAAATCAAGCAATGCTTCTTACTTATCTCT 161788
Qy	707	heGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgT	727
Db	161787	TAGAGCAGATGGGTAAT--TTCCAGCATTCAGACAA-----	CAGTTCCAGAAAC 161740
Qy	727	yrProProAlaSerAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnT	747
Db	161739	AAGCTGAGGAGCTTCATCTGTCGCAAACTAT-----	GGAACCTGGAGGTGACACAGAGC 161686
Qy	747	yrGlnSerGln-----	SerValaSerMetProGlnLeuSerP 759
Db	161685	ACCCACCAACCAATATGCTCTTAACATCTGAAATGATCTTAAACTTGCACACCAACAC	161626

Oy		759	Proteinase	mechE	ProH1Sa	anastTy	SerMetYr	ThrGlnGlySerSera	AnHisP	779
Db		161625	CACCTCCCTAGCGTGGCACCACCACTTCCTGCATATATTATCATGCCAAGTTGGACAACAAGTTTGC							161566
Oy		779	rovalSerGInPrometValGInGInTYrGInProGlnAla---	SerMetProAsnGlnAla	798					
Db		161565	CT---ACCGACAGATGAGTGACGACAAAGTATCAACGACGACGATGTATTATGACTCAAGA							161509
Oy		798	aNTyGCLProlLeProSerTYrGInGlnAlaAsn-PheHisGlyVal-----							813
Db		161508	ATTATGTCATATYGCCACAGATTGACGATTCCTTAATCTTAGCGTACAGCAATTCATACTCC							161449
Oy		814	-----ThrThAsnGlnAlaGlnAnLeuAn							822
Db		161448	CATTCTTCCTTATTCGCCAAGTAACCTTGCCACCAAAATATATAGATAGGGTTTTGG---							161392
Oy		823	ProSerGInPheGlnAlaAlaMetGInProProAlaAspLysAlaAsn-----							838
Db		161391	---GCAGTGTTTCAGAGATCTCTTATGCTTATGCTTAATGATAGGGTAACATATAGGTTTCTT							161335
Oy		839	LeuGInLProGInAsnGlnAlaLeuArgLeuGInProMetLlSerGlyAsp-----							855
Db		161334	TTCACGCGCGCACCATCAACAAATGTGTGGACCTCGTCGACACAAAGCTCCTGTTGGTCGC							161275
Oy		856	-----GlyGInGly							858
Db		161274	CAATTCCCTCCCTGCATATGCAGACTCCTGCTGCCGACCAAGCTCCTGTTGCTGCACAAAGCT							161215
Oy		859	ThrThrAspGlyGluValAspLysAsnGlnArgTYrGInSerThrLeuGInPheAlaAla							878
Db		161214	TCTGCTGAC---GAGCGACGAACGSAACGAAAGTATCAGGCGCACTCTTCAGTTGGCTCAG							161158
Oy		879	AenLeuLeuLeuGlnLlLeGlnGInLysGlnGInGlnSerSerTYrThrProAlaGly							898
Db		161157	GCGCTGTGGGTGCTAGCTCAACAGAAAGCCTGGAATCAACTTAAGGCTATCCTTGAAAT							161098
Oy		899	Gln							899
Db		161097	CMA							161095
RESULT 12										
AK100889			3455 bp	mRNA						linear PLN 24-JUL-2003
LOCUS										
DEFINITION			Oryza sativa (japonica cultivar-group) cDNA clone:0023129A05, full							
ACCESSION										
VERSION			AK100889.1							GI:32986098
KEYWORDS			FLI CDNA; CAP trapper.							
SOURCE			Oryza sativa (japonica cultivar-group)							
ORGANISM			Oryza sativa							(japonica cultivar-group)
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
			Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;							
			Ethnarioidae; Oryzeae; Oryza.							
REFERENCE			1							
AUTHORS			The Rice Full-length cDNA Consortium, National Institute of							
			Kiruchiological Sciences Rice Full-length cDNA Project Team,							
			Kiyochi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,							
			Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,							
			Ohtsuka,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,							
			Chikumi,K., Shishiki,T., Foundation of Advancement of International							
			Science Genome Sequencing & Analysis Group; Otomo,Y., Murakami,K.,							
			Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,							
			Kurosaki,T., Kodama,T., Maeda,H., Kobayashi,M., Xie,Q., Lu,M.,							
			Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niihara,J.,							
			Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,							
			Kusumegi,T., Oka,M., Ryu,R., Ueda,D., Matsubara,K., RIKEN;							
			Kawai,J., Carinici,P., Adachi,T., Aizawa,K., Arakawa,T., Fukuda,S.,							
			Hara,A., Hashidume,W., Hayatsu,N., Imctani,K., Ishii,Y., Itoh,M.,							
			Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,							
			Saito,R., Saeki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,							
			Yoshino,M., and Hayashizaki,Y.							
TITLE			Collection, mapping, and annotation of over 28,000 cDNA clones from							

JOURNAL
Science 301 (5631), 376-379 (2003)

MEDLINE
22752273

PUBMED
12669764

REFERENCE
2 (bases 1 to 3455)

AUTHORS
Aochi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hasegaki,T., Hara,A., Hashizume,W., Hori,F., Hotta,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hirooka,T., Imotani,K., Ishibiki,J., Ishii,Y., Ikeda,R., Imamura,K., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Otsu,N., Oka,Y., Ootomo,Y., Ryu,R., Satoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoch,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Suiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akishira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

TITLE
Direct Submission

JOURNAL
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondo, Tsukuba, Ibaraki 305-8602 Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT
This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project/CDNA/Kikuchi,S.,Satoch,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.

PLS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Ryu,R., Sugano,S., Suiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,C. and Murakami,K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Aochi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hasegaki,T., Hara,A., Hashizume,W., Hayashizaki,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Iida,J., Iida,Y., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Otsu,N., Oka,Y., Satoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sogabe,Y., Satoch,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akishira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES
Source
Location/Qualifiers
1. 3455
/organism="Oryza sativa [japonica cultivar-group]"
/mol_type="mRNA"
/cultiyar="Nipponbare"
/DB_xref="taxon:39947"
/clone="J023129A05"

ORIGIN
Alignment Scores:
Pred. NO.: 2.05e-23 3455
Score: 799.00 298
Percent Similarity: 37.20% Conservative: 119

Best Local Similarity: 26.58% Mismatches: 325
 Query Match: 16.64% Indels: 378
 DB: 0 Gaps: 45

US-09-920-705-3 (1-901) x AK100889 (1-3455)

```

QY 7 ProPheArgAlaAspAspSerGlyPhe-----GlnSerAsnLeuTrpVal 22
DB 203 CCGCGCGGCGCTCTCTCCGGGTGGGCGGACCGCGCTCGCGGACCTGTGGTG 262
QY 23 GlySerLeuThrProGluThrThrGlnSerAspLeuThrGluLeuPheGlyArgTyrGly 42
DB 263 GGCAGCTCTCCCGCGCGCTCGCGCGGACCTCTCGAGGCTCTTCCTCCGCTGGCG 322
QY 43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
DB 323 GACGTGAGGAGGACCTCCCGTACCCCGCGGAGCTTCGCTTCGAGGATTCGCGCG 382
QY 63 ValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyValAlaAsnLeuAsnGlySerGln 82
DB 383 GAGGAGGACCGCTCGCGCGGCTCGCGGAGCTCGAGGAGATCCACTCCGCGGCGCGCC 442
QY 83 IleValIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIle 102
DB 443 ATTAGGATCGAGTTTCCAGGCGGATAAAGATTCAAGTAGC-----484
QY 103 GlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGlu 122
DB 485 -----TCTATGATGAC-----AGATATCTCAACAATGCT 514
QY 123 AspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAsp 142
DB 515 GATCAAGAAGCGTTTACTGAAACGAGA-----541
QY 143 AlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAsp 162
DB 541 -----541
QY 163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
DB 542 -----AGGATATGAGCATTAAGTCTGAAAA-----TCACTGAT 577
QY 183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
DB 578 AAATCCAAAAGACGAGCA-----GCA 601
QY 203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
DB 602 GAACCTAGTAGAATATTATGATAGGATTTCTCTGTTGCTGAGGTA-----GATGAG 655
QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleArgValLysSerTyrPro 242
DB 656 GCAACTCTCTGGGAAACCTTTTCACTTTTGGTAGAGTTGTCAAGATTAATCAATTCCTCA 715
QY 243 SerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlu 262
DB 716 GGGCGTACTATGCTATTTGCGAGTACTACTATTGTCAGCGGAGCAGGCGAGGAGGAGAA 775
QY 263 GlyLeuGlnGlyArgLeuPheAsnProArgIleLysIleMetLysSer-----279
DB 776 ACACCTCAGAGGAATATTTCATATACCTTCGAGTGAATTTGCTTTCTCGAGTGAAC 835
QY 280 -----Asn 280
DB 836 AGTGTTCAGCAGAAATTTGGAAGAGTTCCTAGATGCCCATATTTCCCCCATTTAAAC 895
QY 281 AspGluLeuProPro-----GluGlnAspAspThrSerPheTyrSerGlyMetLys 297
DB 896 TCTAGTGTTAAGCTATATTATTCAGGAGCAGAGATTTTGAAGATTT-----940
QY 298 ArgSerArgThrAspMetPheAsnAsnAspPro-----SerCysValSerSerProHis 315
DB 941 -----CCTAGGCGTAGGCTTTTATAGTCTCTCAAGAGATATGTATACATCATCTCCACAT 997

```

```

QY 316 SerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsn 335
DB 998 TAT-----GGCCCTAAGAGACTTCTTCA-----1021
QY 336 GlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsn 355
DB 1022 -----GATCATGATGATGTGGTTTCAGCAGGATTAATTTATTCGA-----1063
QY 356 GlyThrGlyIleLeuProSerProThrGlyIleLeuProSerPro-----372
DB 1064 -----TATGACCTGAGTAGAGCTCGTCAAGAAAGAGATGTGAGGACCATAT 1159
QY 373 -----AlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
DB 1100 TTTGAACCTTTTAGATATCAAGGCTCGGTCCAGAAAGAGATGTGAGGACCATAT 1159
QY 388 GluGlyTyr-----390
DB 1160 GAACAGCATAGGCGTAGCCCTGCTGTGATGCAACCATGCAACAATTCATTGAGCGA 1219
QY 391 -----AspProAlaGluLeuValArgGlu-----398
DB 1220 TCTCAGGAGGCTTACATTAAGAGATTCGTGATATGCTAGGAAAGATCCATACCATTT 1279
QY 399 SerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGlu 418
DB 1280 TCMAAGAAATGAGAGACTGTGAAAGCACATGACTCTGAATCTTCT--GAATACCTTTTC 1316
QY 419 ArgSerPheGlyArgGlySerValAlaAla-----AspProIleArgGlyPro 434
DB 1337 TCTGAATTTGATCGAGGAAAGTTGGCTCTGCTTAACCAAGAGAGCCCTTCTATGTGTG 1396
QY 435 ProAspSerAsp-----438
DB 1397 CCAGATGATGACATATACACCCAGAGCTATCACTGCTCTATGATGATGAAATCAT 1456
QY 439 -----HisIle-----440
DB 1457 GTTATCTTTAAGAAATCCAACTCACTGTGTAGATGAGCATATATCAGGAGCATGACAG 1516
QY 440 -----440
DB 1517 GACAGCTTTTATAGCATATGAGAGTGAAGAAATCACTCTGATATCATGAACCCCTT 1576
QY 441 -----TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys 454
DB 1577 CTCAGAAGAAATGAAATGGAGTGTATCAATATGCAAAAGGAGGACACCAATTGGCGA 1636
QY 455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValAlaAsnCys 474
DB 1637 GCGGATGCTCTCCCTGTGGAGAGTTCTTAATCTCAATGCTCCGGAATTTTGGATTGC 1696
QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu 494
DB 1697 ACTGCTAGGACAACCTGGAGATGCTCTTAAGACATATATCAAGCTGCGAGAGAGTGG 1756
QY 495 IleValIlePhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
DB 1757 GTGGTGTTTTGTTCAGAAATATGATGCTGACATGCGACCTATATATGATCATGAAT 1816
QY 515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe 534
DB 1817 TACTTGTGTATAGACGCTGACGAGCTTTGTAACTTTGAGAAAGAGACACTATTT 1876
QY 535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
DB 1877 CTGTTCACACCTTCAGACTTCTCTGAACAAGTACTGAGGTTCCAGATAAGTCAGCATAT 1936
QY 555 TyrGlyValValLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGln 574
DB 1937 TCTGAGATCATTTCTGAAGTTTGAAGTGAAGTCAAGATCCAGAAAGTTTCTTCGCCAATCTCCAAA 1996

```

QY 575 Gluser----- 576
 Db 1997 CCGAAACATTTGGAGTCATTTGAACCATGATGTTGCTCTATGAGATCTAGATGCA 2056
 QY 577 -----GlnSerAsnPro-----LeuHisTyMetAspGlnAlaArgAsp----- 589
 Db 2057 TTGAGAGAGATCAACCCACAGATATACAGGCCACTTCTCAGGGTTTCAATTTATTCGGG 2116
 QY 590 ---SerProAlaAsn-----AlaSerHisSerLeuTyProProAlaGluAsnTy 605
 Db 2117 TTGTGCGTGGAGCTGATTAATCCAGCAAGTCACATTTGGTTCGGCTTACAG----- 2170
 QY 606 ILAArgGLValAProGlnHisLeuThrAla----- 615
 Db 2171 TTGGAATGCTCTTATATCTAGAACTGAATTAGCTCATGAAAGCATCCACTGAC 2230
 QY 616 -----AlaSerTyProSerValSerGluProLeuArgIleProAsn--- 629
 Db 2231 TCCACAGAGGAGATGACATGACAGCAGACAGAACACCCAGATGATTTGCCCTCAGAGA 2290
 QY 630 -----AenAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 643
 Db 2291 TGGTCAGATACATTTACATCCAGTCACAGTTCAGTTTCGA-----AATTG 2335
 QY 644 LeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHis 663
 Db 2336 AATTATTTGGCTGAGAGTGGATCCCATACATACATCATGATAGACACAGAGGCGATAC 2395
 QY 664 Gln-----PrometSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGly 681
 Db 2396 TCATTTGCTCTCCTCAGAGAGTACCAAA-----GTGAGTCA----- 2431
 QY 682 LeuTyArgnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAsp 701
 Db 2432 -----TCAGGCTATGACCA----- 2458
 QY 702 AlaSerAsnGlnSerPhe-----GlnGlnTyArgnGlnTyArgnGlnTyArgnGln 719
 Db 2459 GCATCAACACATGCTCTACCTCCCATGCAACCTGTCATACAGACAGGATGATGACCTCA 2518
 QY 720 LeuProProPro---ProSerArgTyProPro----- 729
 Db 2519 CAACCTTCATCTCCCATATATGCTTCACACAGAGCAATTGACAAATGGCCACTCTT 2578
 QY 730 -----AlaSerAsnAsnPro-----AsnTyThrSerGly 739
 Db 2579 CTTCACCAACAAACCAACCAAGAAAGAGCTGTGACAGCTTAAACAAAGATACAGA 2638
 QY 740 MetVal-----HisGlyAsn-----MetGlnTyArgnSerGlnSerValAsn 753
 Db 2639 TTCAATACGGAATCCACATGACATCTTCAATATGATGCACACAGCTCAGGTTCTATC-- 2695
 QY 754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTySerMetTyThrGln 773
 Db 2696 ---CTGTCCAAAACATCATTTGCCACCTGCTCG-----CCA 2728
 QY 774 GlySerSerAsnHisProValSerGlnProMetValGln----- 786
 Db 2729 TCTGCATCAGATTACAGTTTCATGCGCCACACAGTTCAGAGTTCAAGTCCACCAATCCA 2788
 QY 787 -----GlnTyArgnProGlnAlaSerMetProAsnGlnAsnTyArgnGlyProIleProSer 804
 Db 2789 TCCATTAATGATACACCGAATGCTCTTATGCTTCACAAACCTTTGACCCCTTACCCCT 2848
 QY 805 TyArgnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnLeuAsnLeuAsnProSer 824
 Db 2849 ATGATCCTTCGGGA-----AATCAGCTCATTTCTTCAGCTTTG 2890
 QY 825 GlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGluProGlnAsnGln 844
 Db 2891 AGATCGTTCTGCTCTCTCTTCTGAAAGGCTTCACCCCTTAGGCGACACATCAAGT 2950
 QY 845 AlaLeuArgLeuGlnProMetIleSerGlyAspGlyGln-----Gly 858

Db 2951 GCTTACAGCCCACTTCCTCCCTGTGAGCACACAGATCAGCAGAACCGTCTGCT 3010
 QY 859 ThrThrAspGlyGluValaAspLysAsnGlnArgTyArgnSerThrLeuGlnPheAlaAla 878
 Db 3011 CAGAAAGACCATCAGAGAGACCTCCAAAGCCCTTCAAGCAATTCGAAATTCGACAGA 3070
 QY 879 AsnLeuLeuGlnGlnIleGlnGlnTyArgnGlnGlnGlnGlnGlnGlnGlnGlnGln 898
 Db 3071 ACCCTACTTCACAGATACAGAACATCTAA-----CCTGCTGCTG 3112
 QY 899 Gln 899
 Db 3113 CAG 3115
 RESULT 13
 AX420424 3715 bp DNA linear PAT 18-JUN-2002
 LOCUS AX420424
 DEFINITION Sequence 6 from Patent WO0212518.
 ACCESSION AX420424
 VERSION AX420424.1 GI:21524580
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 REFERENCE
 AUTHORS Amesino, R.M., Schomburg, F.M., Michaels, S.D. and Patton, D.
 TITLES Floral induction gene
 JOURNAL Patent: WO 0212518-A 6 14-FEB-2002;
 WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
 FEATURES
 source
 1..3715
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 /note="PPA promoter plus intron"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,786-17 Length: 3715
 Score: 651.50 Matches: 171
 Percent Similarity: 27.238 Conservative: 0
 Best Local Similarity: 27.238 Mismatches: 0
 Query Match: 13.574 Indels: 457
 DB: 6 Gaps: 1
 US-09-920-705-3 (1-901) x AX420424 (1-3715)
 QY 1 MetAlaLeuSerMetIleProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 Db 1832 ATGGCGTTATCATGAAAGCATTCAGAGCCGATGATTCGGTTTCAGTCAAAACATCTT 1891
 QY 21 TrpValGlySerLeuThrProGluTrpThrArgLysSerAspLeuThrGluLeuPheGlyArg 40
 Db 1892 TGGGTGCTGATGCTTAACGCGGAGAGCAGACAGATCATACCGAGTTGTTGGAGAA 1951
 QY 41 TyArgLysAspIleAspArgIleThrValTySerSerArgGlyPheAlaPheIleTyArg 60
 Db 1952 TACGGGATATGATGATGATACCGGTATTTCTTCACAGAGCTTTGCCTTTATATATC 2011
 QY 61 ArgHisValGlnGlnAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 2012 AGACATGTGAGAAAGCAGTCCACAGCAAGAGGCTTTCAAGAGCAAAATTTGAATGGA 2071
 QY 81 SerGlnIleLysIleGluTyArgAlaArgPro----- 90
 Db 2072 AGTCAATTAAGATGAATGACGACGACCGGTTGTTCTTATCATATCTTCGTTGTTTC 2131
 QY 90 ----- 90
 Db 2132 TCTAATTTGATTTGCTTTTGTCAACGATTAATCTTTTTCGAATTCATGTCAGGT 2191

QY 90 ----- 90
 Db 2192 TCACAAACTTGTGATGCTGTTTAGTCCAAAAATTTCTGTGATCTGTTTTTTT 2251
 QY 90 ----- 90
 Db 2252 TTCTCATCGTACAAATCAAAAGTCGAAACCTAGTTTTTTCTATTATAGTCGTAGCTT 2311
 QY 90 ----- 90
 Db 2312 AAGCGAACTCTATCCGATCGAAAGCTCTTTCTCAATTACTTGGTTATATCGAACT 2371
 QY 90 ----- 90
 Db 2372 CGCGCAAGCCAAACACAGAGAGCTCTGCAAAATTTGATGTTAAAGCATATATACTC 2431
 QY 90 ----- 90
 Db 2432 TTAGCGAATGAGCTCTGCAAAAGATATCATCAACATGTTTACTCTCGTTATGCGAA 2491
 QY 90 ----- 90
 Db 2492 GAAGTTTATCCGATTGCTGTTTATCTGCAAACTTGATCCGCTTAAAGCACTTGAT 2551
 QY 90 ----- 90
 Db 2552 TCAGTTCGCTCAGATGTTAAATCTCACAAGCTTGATGATGATTTGCTACTCCA 2611
 QY 90 ----- 90
 Db 2612 CTGCAATTCGCAATGCTGAAAAATTGGAGCGAGCAAACTCTATCATCGGCCAAACAG 2671
 QY 90 ----- 90
 Db 2672 TAAGAAATTTTGAGTTTAAAGTTCAGTTTCTCGCAAAAAATCAACCGGTGAGAAAT 2731
 QY 90 ----- 90
 Db 2732 TTGCTATGCGCGTAGTAGATATCAATACTGCATCTCGGAAACAGAAAAATTTCTGGAA 2791
 QY 90 ----- 90
 Db 2792 CCGTATGCTCTTACTCTCTAGTATGCTGCGACATATCATCATGCTGAGGAGTTT 2851
 QY 90 ----- 90
 Db 2852 GGGCTTGTGCTCGATGCTTCAAGAAAGCAAAATTAGTGGTTGTAAGCGGCTAAC 2911
 QY 90 ----- 90
 Db 2912 TCTACCAATCAGAAAGCTTATTTCAGAAACCATGTTTCTTCATTTCCATCCCTA 2971
 QY 90 ----- 90
 Db 2972 CGTACATCTGGCTTTCCCATTTGTTAGTTCTTTAGGAAGTTGACTTCTTCAACCAATC 3031
 QY 90 ----- 90
 Db 3032 TGGGATTGATACAGTTGAGTGAGCAAGACAAATATTCATTGACGAGCACTCAATTT 3091
 QY 90 ----- 90
 Db 3092 CAATGCTTGGCGGCTTTGAGTACTTAAATGAGAGTTTCAGGATGTTATTGAGGAC 3151
 QY 90 ----- 90
 Db 3152 TCAGGCAATGATAAAGACAGAAAGTTTTCAGAAAGATTAATGTTCTTTCCGCTTGG 3211
 QY 90 ----- 90
 Db 3212 GGAAGTACATCTCTCTACAGACAAAAATCTGATGTTTAAAGCAAGAGAACTATTACA 3271

QY 90 ----- 90
 Db 3272 CTGATTTAATAGAGAAAGAGAGATGTTTCTCTCGGCAATTTATTGTTTAGGA 3331
 QY 90 ----- 90
 Db 3332 AGCAATGATATGAATGSGTGTGATGATGTTGAATTAAGTATGTTGTGTGTT 3391
 QY 90 ----- 90
 Db 3392 AGTTCTTGATGTTTGAATGCTTATTCCTTGCAACTATCGGATGATATCGCTTCT 3451
 QY 91 ----- 103
 Db 3452 ATCGACTTTTCTTTGTCGAGCAAACTGTGTAAGATCTATGAGTGGAGAAATCGGC 3511
 QY 104 ProaenValSerLysAspAspLeuGluGluPheSerLysPheGlyLysIleGluAsp 123
 Db 3512 CTAATGCTCCAGAGATGACCTGGAGAGAGGTTACAGCAAGTTTGGCAAAATCGAGAT 3571
 QY 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspLysTyrGluMetAspAla 143
 Db 3572 TTAGGTTTCTCAGAGAACGAGACAGACAGCTTTCATTTATTTATGATGATGATGATGCT 3631
 QY 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
 Db 3632 TTACAGGCTAAGAGATGAATGAAAGCTTATGGGTGTAGCTTTTGGCGTTGATTTT 3691
 QY 164 LeuArgSerGlnAlaProLysLys 171
 Db 3692 CTCGGTCACAAAGCGCCAAAAAAA 3715

RESULT 14
 AP004011/c
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
 BAC clone:OUJ381_H04.
 AP004011
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, BAC
 clone:OUJ381_H04
 Published Only in Database (2001)
 2 (bases 1 to 130961)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (01-AUG-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program; Kamondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 On Jul 27, 2004 this sequence version replaced gi:46092388.
 Genes were predicted from the integrated results of the following:
 GENSCAN (http://CCR-081.mlt.edu/GENSCAN.html), FGENESH
 (http://www.softberry.com/), GeneMark.hmm
 (http://opal.biology.gatech.edu/GeneMark/), GLIMMER
 (http://www.tigr.org/cdb/glimmer/glimmer3.html), RICEHMM
 (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
 (http://biointformatics.iastate.edu/cgi-bin/bp.cgi), sIm4
 (http://globin.cse.psu.edu/html/docs/sIm4.html), gap2
 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
 genomic sequence was searched against NCBI Nonredundant Protein
 database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
 sequence database at RGP or DDBJ. Protein homologues of the coding
 regions were searched against NCBI Nonredundant Protein database
 with BLASTP. ESTs represent the identified cDNA sequences using

BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to RGSF standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OJ1381_H04 clone has an overlap with P0592C05 (DDBJ: AF004756) clone at 5' end and an overlap with OSUNBA0062409 (DDBJ: AF004736) clone at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

```
source
1. .130961
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="genomic DNA"
   /cultivar="Nipponbare"
   /db_xref="taxon:39947"
   /chromosome="9"
   /clone="OJ1381_H04"
   /complement(3074..6118)
   /gene="OJ1381_H04.1"
   /complement(join(3074..3542,3637..3762,4173..4312,5566..6063))
   /gene="OJ1381_H04.1"
   /note="supported by full-length cDNA(s) : AK104965"
   /complement(join(3105..3542,3637..3762,4173..4312,5566..6118))
   /gene="OJ1381_H04.1"
   /note="supported by full-length cDNA(s) : AK073729"
   /complement(join(3402..3542,3637..3762,4173..4312,5566..6005))
   /gene="OJ1381_H04.1"
   /note="contains EST(s) : AU174422(S6559), C25488(S6559) contains full-length cDNA(s) : AK104965, AK073729"
   /codon_start=1
   /product="putative ubiquitin-conjugating enzyme family protein"
   /protein_id="BAD32975.1"
   /db_xref="GI:50725506"
   /translation="MTLSSGAGSSVVRNRLLELRGEKIGDGVSYGMDDAD
DIYRSGTITIGPHNSVHEGRIVQLKFCDDYPEKPSVRPHSRINNSCVNHETGA
VSRKRGVLANQRETYMETITLTOLKKNATPDQNRKLVOPEGTFF"
   /complement(8645..9266)
   /gene="OJ1381_H04.2"
   /complement(8645..9266)
   /gene="OJ1381_H04.2"
   /note="probably inactive due to 3'exon missing in CDS
pseudogene, serine/threonine phosphatase PP7-related"
   /pseudo
   /complement(9412..10109)
   /gene="OJ1381_H04.3"
   /complement(9412..10109)
   /gene="OJ1381_H04.3"
   /note="probably inactive due to 3'exon missing in CDS
pseudogene, transposase related protein"
   /pseudo
   /complement(14903..15292)
   /gene="OJ1381_H04.4"
   /complement(<14903..>15292)
   /gene="OJ1381_H04.4"
   /note="start and end point are not identified"
```

```
CDS
complement(14903..15292)
/gene="OJ1381_H04.4"
/note="similar to Oryza sativa chromosome 7, P0669H03.22"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD32976.1"
/db_xref="GI:50725507"
/translation="MASRWVATARTKTSRGCDRGORWRRLSPPTRMASRIIAAAC
SSPSVDVQDCSASCYRMRRLSLPWIQPMISATNAVEAFTEVDDGSD
SGCRQWRERKRKRKEGDEDT"
join(16127..16171,16261..16372,16416..16460,16533..16753)
join(16127..16171,16261..16372,16416..16460,16533..16753)
join(16127..16171,16261..16372,16416..16460,16533..16753)
/note="hypothetical ORF
predicted by GENSCAN
this category is not included in RGSF standard"
16879..18432
/gene="OJ1381_H04.6"
<16879..>18432
/gene="OJ1381_H04.6"
/note="start and end point are not identified"
16879..18432
/gene="OJ1381_H04.6"
/codon_start=1
/product="putative monosaccharide transporter 6"
/protein_id="BAD32977.1"
/db_xref="GI:50725508"
/translation="VVAASGAGRPYPGGLTFVSMACLVAAATGGLIFGYDIGSGV
TSMDFPLSRFPSPVYRAASAAAGAGQYCFPDSGLTFMPSISYLAALASLGAT
VTRVAGRRKMSMPAGGLVFLAGCALGAANVAMLVGRYLIVGIGFANQSVVYLS
MAPARMGILNNGFOMITTYGLANLINYCARFAGMGWRLSLAALVPAVATAG
ALFDPETNSILERRRGRARMQVRGEGYDMDENDVLAAGASIAVASPRDI
LKRNRPLVAVAIPLFOQLTGIVNIFVAVLRTIFPGSGLMSAVITGGA
ATLVAVLAVRVRGALPLRGAQVNAQAAIGARLFGSGTAAPAGVAAVMA
AMCVVAAPFAMSGPLAMLVSEVWFLVRPAGQITVAVNAAVFAVAQATPLICR
LRFVFPFAGVAAVMTAFVALLFVPETKGVPIEDMAAVMSDHWYKRFVDDGDGAR
RGDIEMGHK"
complement(20105..23575)
/gene="OJ1381_H04.7"
complement(join(20105..20513,20597..20726,20898..20966,
21052..21171,21461..21545,21670..21755,21867..22019,
22113..22235,23079..23575))
/gene="OJ1381_H04.7"
/note="supported by full-length cDNA(s) : AK071036"
complement(join(20437..20513,20597..20726,20898..20966,
21052..21171,21461..21545,21670..21755,21867..22019,
22113..22235,23079..23462))
/gene="OJ1381_H04.7"
/note="contains EST(s) : AU100795(C51066), C27141(C51066)
contains full-length cDNA(s) : AK071036"
/codon_start=1
/product="phosphoenolpyruvate/phosphate translocator"
/protein_id="BAD32978.1"
/db_xref="GI:50725509"
/translation="MQSAAYAGLAPPCGATTAAPLQLRNPSRPGVCGPQLPR
GLRSLAVPRAGISAKRTGLVPASBPQEDERRRGARDAVARTAAAGABEGGGL
AKTLQLGALFGILWYLFNTVFNINQVAKVAPPYLINTVQAVGTVALFLWITGIL
KRPKISGAOLAIPLAVWITGNLFTMNSLQKAVSFTHITIKAPRPSVLSALFL
GEMPPPVVLSLVPVIGVALASLLEAFENAGFVSAASAVTFQSRNVLSSKLAWKC
BSLIDNITLFSIITVMSFPLAPVTLTLEGVGVPTVTAOSGLNKKOYTRSLIAFC
PHAYQVSYMLIARSPYTHSVGNCKRVVIVTSVLFPRTPVSINSLGTVALAGV
FLYSQRLKRPKRTA"
complement(24740..25376)
/gene="OJ1381_H04.8"
complement(join(<24740..24897,25172..>25376))
/gene="OJ1381_H04.8"
/note="start and end point are not identified"
complement(join(24740..24897,25172..25376))
/gene="OJ1381_H04.8"
/note="predicted by Genemark.hmm etc."
/codon_start=1
/product="hypothetical protein"
```

```

/protein_id="BAD3297.1"
/db_xref="GI:5072510"
/translation="MEVSRGSPDNPSPLOQDATSLRRODAEREREEDLPEPPCHC
TREMLRGKARKATTPRRRLAEVTTATLIDLNLCSGGVSGAAGVATMEBEENV
EEVWIGDKVIMLSVPS"
/complement(26622..27293)
/gene="OJ1381_H04.9"
/complement(join(26622..26706,26948..27094,27220..>27293))
/gene="OJ1381_H04.9"
/feature="start and end point are not identified"
/complement(join(26622..26706,26948..27094,27220..27293))
/feature="OJ1381_H04.9"
/feature="predicted by GeneMark.hmm etc."
/codon_start=1
/product="hypothetical protein"

Alignment Scores:
Pred. No.:      8.31e-13      Length:      130961
Score:          576.00      Matches:      257
Percent Similarity: 31.28%      Conservative: 99
Best Local Similarity: 22.58%      Mismatches: 312
Query Match:    12.00%      Indels:      470
DB:              8          Gaps:          41

US-09-920-705-3 (1-901) x AP004011 (1-130961)

QY 144 LeuGlnAlaIysSerMetAsnGlyLysProMetGlyLysSerPheLeuArgValAspPhe 163
Db 95385 CTGGACCTGAATATTTGAATTATTCAGATAAAGGTTCAAGCTCTATGATGACAA 95326

QY 164 LeuArgSerGlnAlaProLysLysGluGlnTTPAlaGlySerTyrAspAsnArgAsnGly 183
Db 95325 TACTCAACAATCTGTATCAAGACGT-----TTTCTGAACGAGGAGAG 95281

QY 184 AsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGln 203
Db 95280 AATCAGAGATCAAGCTCTGTAATAATCACTGATTAATCCAAAAGAGCCGACGACAA 95221

QY 204 ProSerLysValLeuTTPAlaGlyPheProProthAlaThrGlnCysAsnAspGluGln 223
Db 95220 CCTAGTAGAGTATATGATAGAGTTTCTGTTGATGTAAGGTA-----GATGAGCA 95167

QY 224 IleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSer 243
Db 95166 ACTCTCTGGGAACCTTTTCACTTTGGTGAAGTTGTCAAGATTAATCAATTTCCAGAG 95107

QY 244 ArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGly 263
Db 95106 CGTACTTATGACTTTGTCCAGTACACTACTATTGACGCGCATGACGCGCAAGGAAACA 95047

QY 264 LeuGlnGlyArgLeuPheAsnAsnProAlaGlyIleLysIleMetTyrSer----- 279
Db 95046 CTGCAGGGAATATTTTCAATTAACCTCGATTGACATTTGCTTTTCGAGTAGACAGT 94987

QY 280 -----Asp 281
Db 94986 GTTTACAGCAGATTTGAAAAGTTCTTAGTACGCCCATATTTCCCCCATTTAAACCT 94927

QY 282 GluLeuProPro-----GluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
Db 94926 AGTGTAGACTTATATTCAGAGGACAGATTTTGAAGATTT----- 94885

QY 299 SerArgThrAspMetPheAsnAsnAspPro-----SerCysValSerSerProHisSer 316
Db 94884 CCTAGGCGTACGCTTTTGTATGATGCTCCAGAGATATGATACATTCCTCCACATTAAT 94825

QY 317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
Db 94824 -----GGCCCTTAAGAGACTTCTAGA----- 94804

QY 337 AlaGluTyrAsnAspValValGlyLysGluProAsnTTPArgArgProSerAlaAsnGly 356
Db 94803 ---GATCAATGATGATGGGTTTCAGCAGGAGGATATATATTTCGA----- 94762

```

```

QY 357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerPro----- 372
Db 94761 -----TATGAGCTGAGTAGAGAGCCCTGATCTTATGATTAATTT 94723

QY 373 -----AlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
Db 94722 GAACCTTTTAGATACAAAGGCTCGGTCCAGAAAGAGATGTCTGAGGACCCATATGAA 94663

QY 389 GlyTyr----- 390
Db 94662 CAGCATAGCCGTAAGCCCTGCTGTATGATGACCATGCGCAAAACATTTCCATTCAGCGATCT 94603

QY 391 -----AspProAlaGlnLeuValArgGlu-----Ser 399
Db 94602 CAGGAGCTTACATTCATTAAGGATTCCTCGTATGCTAGGAAAGATCCATCCATTTTCA 94543

QY 400 LysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg 419
Db 94542 AAGAAAGTTGAGACTGTGGAAGCACATGACTGAACTTCTCT---GAATACCTTTCTCT 94486

QY 420 SerPheGlyArgGlySerValAlaAla-----ArgProIleArgGlyProPro 435
Db 94485 GAATTTGATCGAGGAAAGTTGGCTGTGCTCCATCCCAAGAGGCCCTTCTATGTGTGTCCA 94426

QY 436 AspSerAsp----- 438
Db 94425 GATGATGATCATACACCCAGAGCTATCACTGCTCTATGATGATGAGAAATCATGTT 94366

QY 439 -----HisIle----- 440
Db 94365 GATCCTTTAAGGATCCAACTCCACTGTGATGATGATGATGATGATGATGATGATGATGAT 94306

QY 440 ----- 440
Db 94305 AGCTTTTCTAGCATGTAGAAAGTGAAGATCAACTCTGATATACATGAACCCCTTCTC 94246

QY 441 -----TTPArgGlyMetIleAlaIleGlyLysIleThrProValCysGlyAla 455
Db 94245 AAGAAAGTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 94186

QY 456 ArgCysValProMetGlyLysGlyIleGluThrLysLeu----- 468
Db 94185 CGATGCTTCCCTGTGGGAAGTTCTTAATCTGATGCTGTAAGTCTGGGTTCTTCACTGT 94126

QY 468 ----- 468
Db 94125 ATTTCTTTGTTCACTTGAATATCATTTATGACATCTACAAATATAGCTCTGTATCGT 94066

QY 468 ----- 468
Db 94065 CTTAATAAATAACAATAATTTCTCTTCATGCTTCAAGATATATATATATATATATATATATAT 94006

QY 468 ----- 468
Db 94005 TCATATATATCGTTTGTCTTACATTAATTTATGATGATGATGATGATGATGATGATGATGAT 93946

QY 468 ----- 468
Db 93945 ATGAAGCTTTTGAAGCTATGTTTGTTCAGCTGTTCTTGAATTAATTAATTAATCTCTG 93886

QY 468 ----- 468
Db 93885 CACCATTTGCTCTACTCATTAATCTTAATAATAATAAGAGCTTTGCTTCAAGGTCCAAAC 93826

QY 468 ----- 468
Db 93825 TACATATCTAGTTGATCAATTTCTATCTGATCTTTTGTGTATCTGATGATGATGATGATGATGAT 93766

QY 468 ----- 468
Db 93765 TTGTGTTTATGCTCTGTCTTATTTGTTCTTATGATTTGACATCTCTAAAGTTCAACATTTA 93706

```

QY 469 -----ProgluValValaenCysSerAlaArg 477
 Db 93705 GTTTTAAACGGCTGTTGATTTTCACTTTCAGGCGGAAATTTTGGATTGCTACCTGCGAGG 93646
 QY 478 ThrAspLeuAsnMetLeuAlaValHisIleTyrAlaValAlaIleGlyCysGluIleValPhe 497
 Db 93645 ACAAGCCTGGAGATGCTCTGTAAGCACTATTACCAACCTGCCAGACAGCTGGTGGTGT 93586
 QY 498 PheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuAlaGlyTyrLeuSer 517
 Db 93585 TTTGTTCCAGAAATGATGCTGATGCGATGCGATTAATGATTAATCAAGAAATTAACCTTGGT 93526
 QY 518 SerIleAspArgAlaGluValAlaValLeuAspAspGlyThrThrLeuPheLeuValPro 537
 Db 93525 GATAGAGAGCTGTCAGAGATTTGTAACCTTGAGAAAGAGAGACAGCTTATTTCTTTGTTCA 93466
 QY 538 ProSerAspPheLeuThrAspValLeuGluValThrArgGluGluArgLeuTyrGlyVal 557
 Db 93465 CCTCAGACTTCTCTGAAACAAGTACTGAGGGTTCCAGAGTAAGTCAAGCATATCTGGAGTC 93406
 QY 558 ValLeuIleLeuProProProAlaValProValThrAlaSerTyrArgGluGluSer 576
 Db 93405 ATTCTGAAGTTTGAGAGATCCAGAGATTCAGAGATTTCTGCGCAACTGCGCAACAGAAACA 93346
 QY 577 -----Gln 577
 Db 93345 TTTGTAGTCATTTGAACCATGATGTTCTGCTCATAGAGATCTAAGTCATTGAGAGAGA 93286
 QY 578 SerAsnPro-----LeuHisTyrMetAspGluAlaArgAsp-----SerPro 591
 Db 93285 ATCAACCCACAGATATCAAGCACTTCTCAGGGTTCAAGATTAATCTCGGGTTGCGCT 93226
 QY 592 AlaAsn-----AlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgGly 608
 Db 93225 GGAAGCTAATATCCAGATGACATTTGTTCCGCTTAACAG-----TTTGAAAT 93172
 QY 609 AlaProGluHisLeuThrAla----- 615
 Db 93171 GCTCCTTCATATCTGAATCTGAATTAGCTCATCAAAAGATCCACTGCCACAGG 93112
 QY 616 -----AlaSerIleProSerValSerGluProLeuArgIleProAsn----- 629
 Db 93111 GAGATAGCATGACAGACAGACAGCAACACCAAGATTAATGCGCTCAAGATGTCAGAT 93052
 QY 630 -----AsnAlaIleProGluAlaGlyValSerLeuThrProGluLeuAlaThr 646
 Db 93051 AACATTACATCCAGTCCAGGTTCCGA-----AATTGAATTAATTG 93007
 QY 647 LeuAlaSerIleLeuProAlaThrSerGluProAlaAlaProGluSerHisGln----- 664
 Db 93006 GCTGAGAGTCCGATCCCATCACTACATCACTGATAGACACACAGAGGATCACTCATTTGCT 92947
 QY 665 ProMetSerGlyProSerThrValSerThrAlaHisGlnSerAsnGlyLeuTyrAsn 684
 Db 92946 CCTCAAGAGATCAAAA-----GTGAGTACA-----TCA 92917
 QY 685 GlyGluAlaProSerGluAlaTyrPheArgGlyProGluThrValHisAspAlaSerAsn 704
 Db 92916 GGGTATGACCA-----GTTCCAGATGAGCATCAAC 92884
 QY 705 GlnSerPhe-----GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGluLeuProPro 722
 Db 92883 ATGTCTTACCTTCCATGCACTGCACTGATCAGACAGAGTACTGAGACTCAACACCTCCA 92824
 QY 723 Pro-----ProSerArgTyrProPro-----Ala 730
 Db 92823 TCTCTCCATTAATTCGTTCCACAGAGCACTTGCACAATTGCGACTTTTGACAA 92764
 QY 731 SerAsnAsnPro-----AsnTyrThrSerGlyMetVal--- 741
 Db 92763 CAAAACCAACAGAGAAAGAGCTGTGACAGCTTAAACAAAGATTCAGATTCATACGG 92704
 QY 742 -----HisGlyAsn-----MetGlnTyrGlnSerGlnSerValAsnMetProGln 756

Db 92703 AATCCACATGAGCATTTCTTAATGATGCCACAGCTCAGGTTCTATC-----CCTGTC 92650
 QY 757 LeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnIleSer 776
 Db 92649 CAAAACCTATGGCACCTGCTCG-----CCATCGCATCA 92614
 QY 777 AsnHisProValSerGlnProMetValGln-----Gln 787
 Db 92613 CAGTTAAGGTTTATGCGGCACAGTTCAAGGTTTACGTGSCAACAAATCCATTCATTAG 92554
 QY 788 TyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGln 807
 Db 92553 CATACACCAAGTCTCTTATGCTTCTCAACACATTTGCTTACCCCTTAATGATCCT 92494
 QY 808 AlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGln 827
 Db 92493 TCGGGA-----AATCCAGCTCATTTCCATGCTTGAAGATGCTTC 92452
 QY 828 AlaAlaMetGlnProProAlaAspValAlaAsnLeuGluProGlnAsnGlnAlaLeuArg 847
 Db 92451 GTCCCTCTCTTCTTAAGGCTTCCACCCCTTAGGACACACATCAAGTCTTAACA 92392
 QY 848 LeuGlnProMetIleSerGlyAspGlyGln-----GlyThrThrAsp 861
 Db 92391 GCGCAACCTGCACTTCCCTGACACAGACTAGCCACAGCAACCTCTGCTCAGAAAGC 92332
 QY 862 GlyGluValAspIleAsnGlnIleArgTyrGlnSerThrLeuGlnPheAlaAsnLeuLeu 881
 Db 92331 CATCAGAGAGACCCCTCAAAAGCGCTTCAAGCAACATTCATTAATGCGCAACCACTTACT 92272
 QY 882 LeuGlnIleGlnGlnIleAsnGlnGlnGlnGlnIleAsnSerGlyThrProAlaGlyGln 899
 Db 92271 CAGCAGATACAGCAACATCTAAA-----CTGCTGCGCAG 92236
 RESULT 15
 ATTPI17
 LOCUS
 DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone TIP17 (BSSA project).
 ACCESSION AL049730
 VERSION AL049730.2 GI:5823567
 KEYWORDS
 ORGANISM Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 REFERENCE EU Arabidopsis sequencing project.
 AUTHORS Direct Submision
 JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemke@mip.biochem.mpg.de, mayew@ip.biochem.mpg.de
 COMMENT On Sep 5, 1999 this sequence version replaced gi:4725940. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
 FEATURES
 source
 1..137519
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 1..18855
 /note="overlap to BAC T4C9; please refer to this entry for analysis and annotation"
 misc_feature
 17536..20463
 /gene="AT4G12420"
 /complement(join(17536..17635,17716..17915,18006..18084,18165..18530,18612..18978,19061..19199,19285..19555,19663..19772,20332..20463))

exon complement (33880, .34038)
/gene="AT4g12430"
/number=6
intron complement (34039, .34111)
/gene="AT4g12430"
/number=6
exon complement (34112, .34219)
/gene="AT4g12430"
/number=7
intron complement (34220, .34294)
/gene="AT4g12430"
/number=7
exon complement (34295, .34366)
/gene="AT4g12430"
/number=8
intron complement (34367, .34468)
/gene="AT4g12430"
/number=8

Alignment Scores:

Pred. No.: 1,1e-12 length: 137519
Score: 573.50 Matches: 240
Percent Similarity: 35.37% Conservative: 90
Best Local Similarity: 25.72% Mismatches: 274
Query Match: 11.95% Indels: 330
Gaps: 36

US-09-920-705-3 (1-901) x ATT1P17 (1-137519)

QY 145 GlnAlaIySerMetLeuAnglyLySPrometGlyGlySerPheLeuArgValAspPheLeu 164
Db 132829 CAGGAAAAGTCATCAACTGATGATCAGCATGAT-----GATATATAT 132870
QY 165 ArgSerGlnAlaProLySlySgIuGIntPrAlaGlySerTryAspAsnArgAnglyAsn 184
Db 132871 CGCCATGATGACAGCGGCTGACAGCAAGAGGCTGCTTTGTCCAAAGGAGCTCAGA 132930
QY 185 MetAsnHiLySerProGIntYrProHiLySerTryGluAspPheLyS-----GlyAsp 201
Db 132931 ATGCGTTAT-----GAGAACCCAGACACATATATGCAATCAAAATGAAATGATAGAAAT 132984
QY 202 ValGlnProSerLyValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAsp 221
Db 132985 GCAGAAACCCAGTACTACTATATATAGGCTTCTCT-----CCTTCGTTGAAAGTATGAT 133038
QY 222 GlnGlnIleLeuHiSaSnAlaMetIleLeuPheGlyGluIleGluArgValLySerTry 241
Db 133039 GATGCGCTCTTGAGGAACGCTTTCTTCTGTTGGAAATTAACAAGGTCAGGTAATC 133098
QY 242 ProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLyS 261
Db 133099 CCGGTCGCGAGTATGATTTGTTCAATTCGAAATCTGATGCGAGCTTGTAAGGGGAAA 133158
QY 262 GlnGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLySileMetTrySerAsnAsp 281
Db 133159 GAAAGCTTCAGGAAAGTTATTTGGCAATCTCGATGTCATATTTGTTTGGAAAGAGT 133218
QY 282 Glu-----LeuProGlu 286
Db 133219 GAACCTTCTCATCTGCGACATGGAAGGGCTCGGCGGTAGATCACGCTCTCCACCTTAC 133278
QY 287 Gln-----AspAsnThrSerPheTryrSerGlyMetLyAspArgSerArgThr---AspMet 303
Db 133279 AGATCTGTGTATGACTCGGATCTTCAGAAAGTTATCTTCAGATGAAACTATGGAAGC 133338
QY 304 PheAsnAsnAspProSerCysValSerSerProHis----- 315
Db 133339 ATACAGCAAGATTCCTCACT---GTGAGGAAACCATATACATTAAGATAGGATTTAGAA 133395
QY 316 -----SerThrGlyIleProGly 321
Db 133396 GATTCTGAAGCTTACATTTTAAACAGGAACGAGACTCAAGTACGATGAGGATCTCTGCA 133455

QY 322 SerMetArgProLeuArgGlyThrAsnGluArgSerTryrAsnGlyValaGluTryrAsnAsp 341
Db 133456 TATGGAAGCTCG---AGATCTACGCTATGATTTCTCAGAGATATGATTAATATCATGCT 133512
QY 342 -ValValGlyLySgIuProAsnTrp-----ArgArgProSerAl 354
Db 133513 ACTCTCGGGGAAA-----TGGGTACTTCATTCGCGATATATCCTCAGAGATTCAG 133563
QY 354 abAnglyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaI 374
Db 133564 ACNAGGA-----GTTCAAGATATGAAAGCCATGGAGCTTACCCAGAAATGACTAC 133614
QY 374 nGlyThrArg-----ArgProMetArgSerAsnProAspSerTrpGluTryrAspPr 392
Db 133615 TACTACCAAGGAAATCAAGAGATTGAAGACAGATCTCGACGCTGAGAGGAGCTTC-CC 133673
QY 392 oAla-----GlnLeu-----ValArgGlySerLySArgThrArgArgAspGlySerValAs 409
Db 133674 AGGGCATCAGCTTTCTCGATATGAGCAAGAAAGAGCCCTTTCAAGGGCATCTGCTGA 133723
QY 409 pGlyPheThrPrometGlyValAspGluArgSerPheGlyArgGly----- 424
Db 133734 T---TTTCCCAAAAGATGATTCATTCAGCGGAACTATGAGGCTGACAGCTAAGATCAA 133790
QY 424 ----- 424
Db 133791 CCAACAGTTGACGACCGCTTAAACCTGAGCTTAAAGAAATGGGACAAAGTACCTTACG 133850
QY 425 -----SerValAlaAlaArgProI 431
Db 133851 AGAACCATGAAAGATTAAATGAGAGATTATCTCTGCCATCAGTTGTCCTGAAAGAAA 133910
QY 431 eArgGlyProProAsp-----SerAspHisIleTryrArgGlyMetI 445
Db 133911 AAGGTACACCCCTGAACTGATGACACATCACTAAAGAGCTGAACTGGGAAGGACTAT 133970
QY 445 eAlaLySgIyGlyThrProValCysCysAlaArgCysValPrometGlyLySgIyIleG 465
Db 133971 TCGTAAAGGAGGCAATCCCATTTGTCAGCAAAATGCTTTCTGTGGGAAAGTGTGGA 134030
QY 465 uThrLySLeu----- 468
Db 134031 CATGATGCTGTAAGTACGCTTAAGCTTATTTCTGTATTTCTGTTGATTTGGCTC 134090
QY 468 ----- 468
Db 134091 TCTATATTGCTTAAGTATTCAATTCACCAAGTATAGTCAATCCGTTGTGATTTGTT 134150
QY 468 ----- 468
Db 134151 TAAGACTTTGCTTAATTTCATTTGCTTTTGTGTATGCTTTAGACTTGTATCAAA 134210
QY 469 -----ProGluValAlaAsnCysSerAl 476
Db 134211 TTAATTTTATGTTTGTATGTTATATAAGCAGGCTGAGTTTCTAGATTTGACGCGC 134270
QY 476 aArgThrAspLeuAsnMetLeuAlaLySileTryrAlaValAlaIleGlyCysGluIleVa 496
Db 134271 AAGAACTGGTTTAAAGCATGCTGCTAAGCATTAACAATCATCTAAAGCATGGCTGCT 134330
QY 496 lPhePheValProAspArgGluGluAspPheAlaSerTryrThrGluPheLeuArgTryrle 516
Db 134331 TTTCTTTGTTCTGGAAGTATGCTGATATTCGTTTATATGATGATTTATGATATATCT 134390
QY 516 uSerSerLySAspArgAlaGlyValAlaLySLeuAspArgGlyThrThrLeuPheLeuVa 536
Db 134391 GGAGAGAGAGCAACGGGAGCGCTTTCTTAATTTGATGACACAACAAGCTTTCTGCTGCT 134450
QY 536 lProProSerAspPheLeuThrAspValleuGlnValThrArgGlnGluArgLeuTryrG 556
Db 134451 GCTTCATCTGATTTCTCCGAGAAAGTACTTAAGTTCTCTGGGAAACTAAGACTCTCTGG 134510
QY 556 yValValleuLySLeuProProAlaValAlaProValThrAlaSerTryrArgGlnGlySe 576

```
Db 134511 AGTATTCTACGTTTA----- 134526
QY 576 rGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHis 596
Db 134527 -----GAAATGGTGGTCTGG 134543
QY 596 sSerLeuTyrProProArgGlnuLeuTyrIleArgGlyAlaProGlnuHisLeuThrAlaAl 616
Db 134544 CTCGGCGGGGCGCTGTTCAACAGCAAGGTGAGAGAAAAGATACAGACTGCTTAATTATTA 134603
QY 616 aSerLysProSerValSerGlnProLeuArgIleProAsnAsnAlaAlaProGlnAlaG 636
Db 134604 TGGTGAACAACATCATTCACAAA-----GCTAGTGAAGCGTTCTCGATGTGGG 134651
QY 636 yValSerLeuThrProGlnuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerG 656
Db 134652 A-----AA 134654
QY 656 nProAlaAlaProGlnuSerHisGlnProMetSerGlyProSerThrValValSerThrAl 676
Db 134655 CCCACGATTCG-----GGTCCACACGCTTTCTGGGTAGTGC 134693
QY 676 aHisGlnSerAsnGlyLeuTyrAsnGlyAlaProSerGlnAlaIleTyrLysArgGlyPr 696
Db 134694 GGGTCGTGACAA-----CAAAAGTCATCAATGATCC 134726
QY 696 oGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrPr 716
Db 134727 ATAC-----GTGAGAGACACAGATATCATGTTGAGTCAACGGTATTCGGA----- 134772
QY 716 oAlaGlyGlnLeuProProProProSerArgTyrProProAlaSerAsnAsnProAsnTy 736
Db 134773 -----TCGAGCTGGCGCCTCGTGATACAAACCCGAGAAG 134807
QY 736 rThr-----SerGlyMetValHisGlyAsnMetG 746
Db 134808 TTCACCTTCATAGATCATCTGTCAAAACACAGCGGGTGTCTCGGGAACACAGCA 134867
QY 746 nTyrGlnSerGlnSerValAsnMetProGlnuLeuSerProLeuProAsnMetProHisAs 766
Db 134868 A-----AATGC 134873
QY 766 nAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValG 786
Db 134874 AGATCTAAGTCGTAT-----CATGATACGAAATCCAGTCCAGC 134915
QY 786 nGlnTyrGlnProGln-----AlaSerMetProAsnGlnAsnTyrG 800
Db 134916 AGGGTTTCAGCTGAACAGCTTACACATTAGCTTCCTGCTACCCAGACAGCAA----- 134970
QY 800 yProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAs 820
Db 134971 -----CAACAGGTACAG-----AACACTCCAAATCAGCTGAGAG 135005
QY 820 nLeuAsnPro-----SerGlnPheGlnAlaAlaMetGlnProProAl 834
Db 135006 AATATGCACCAAGAGCCGAGCAAGTTTCAGCATTTGCAACATGCG-----CAGACACCAAG 135062
QY 834 aAspLysAlaAsnLeuGlnProGlnuAsnGlnAlaLeuArgLeuGlnProMetIleSerG 854
Db 135063 CATTAAGCTCAATGGTAACTCCCAAAATCAAAATGATACAGATACAA-----AGTAG 135113
QY 854 yAspGlyGlnGlyThrThrAspGlyGlyValAspLysAsnGlnArgTyrGlnSerThrIle 874
Db 135114 CAACAGCCCAACAAGAGAAAGAGAGCTAATCCACAGAAAGCGTCTTCAAGCAACATT 135173
QY 874 uGlnPheAlaAlaAsnLeuLeuGlnIleGlnGln 886
Db 135174 ACAGCTAGCAGCTGACATTCTCCAGAGATTCACACA 135210
```

Job time : 10611 secs

PN WO200212518-A2.
 XX 14-FEB-2002.
 PD
 XX
 PF 02-AUG-2001; 2001WO-US024427.
 XX
 XX 03-AUG-2000; 2000US-0222550P.
 PR
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Amesino RM, Schomburg FM, Michaels SD, Patton D;
 XX
 DR WPI: 2002-227160/28.
 DR P-PSDB; ABB07659.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 PT
 XX

PS Claim 5; Page 29-34; 39pp; English.

CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, *frigida* (FRI) and FLC which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA protein encoding cDNA
 XX

SQ Sequence 2706 BP; 779 A; 628 C; 617 G; 682 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3,74e-246	Length:	2706
Score:	4801.00	Matches:	901
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-920-705-3 (1-901) x ABA95167 (1-2706)

QY 1 MetAlaLeuSerMetIysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 Db 1 ATGGCTTATCTATGAGCCATTCAGAGCGATGATTCGCGTTTCAGCAAAACATCTT 60
 QY 21 TrpValGlySerLeuThrProGluThrThrgluserAspLeuThrGluLeuPheGlyArg 40
 Db 61 TGGGTGCTAGAGCTTAACGCCGAGACGACAGAGTCAGATCTGACCAAGTTGTTGAAGA 120
 QY 41 TyrGlyAspGlnLeuAspArgGlyLeuThrValTyrSerSerArgGlyPheAlaPheIleTyr 60
 Db 121 TACGGGATATTGATGATGATCAACGGTGTATCTTACAGAGGCTTTCCTTATATATCTAC 180
 QY 61 ArgHisValGluGluAlaValAlaAlaIleValysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 181 ACAACATGTGAGAGAGAGAGCTGCCACCAAGAGGCTCTTCAAGAGCAAAATTGAAATGA 240
 QY 81 SerGlnIleValysIleGluThrAlaArgProAlaIysProCysIysSerLeuTrpValGly 100
 Db 241 AGTCAAAATTAAATGATGATGATGACGACGACGCAAAACCTTGTAAAGATCTATGGTGGGT 300
 QY 101 GlyIleGlyProAsnValSerIysAspAspLeuGluGluIleuPheSerIysPheGlyLys 120
 Db 301 GGAAATCGGCTTAATGTCTCCAAAGGATGACCTGGAGGAAGATTCAAGCAAGTTTGGGAAA 360

QY 121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
 Db 361 ATCGAGGATTTTATGTTCTTCAGAGACCAAGACAGCTTCATGATTTATATGATG 420
 QY 141 AspAspAlaLeuGlnAlaIysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
 Db 421 GATGATGCTTTACAGGCTTAAGAGCATGATGAAAGAGCTTAATGGGTGATGCTTTGGGT 480
 QY 161 ValAspPheLeuArgSerGlnAlaProLysIysGluGlnIleThrAlaGlySerTyrAspAsn 180
 Db 481 GTTGAATTTTCTCCGCTCACAGCGCCAAAAGAAACATGGCGCTCTTACGATAC 540
 QY 181 ArgAsnGlyAspMetAsnHisIysProGlnTyrProHisSerTyrGluAspPheLysGly 200
 Db 541 AGAAATGGCAATATATCATATTAACCGCATATCTCTCATATGAAAGCTTTAAAGA 600
 QY 201 AspValGlnProSerLysValLeuThrIleGlyPheProProThrAlaThrGlnCysAsn 220
 Db 601 GATGTCAGCCAGATGAGGTTCTGTGATGGGTTCCTCTACTGCTACCAATGCAAT 660
 QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
 Db 661 GATGAGCAAAATTTCTSCAATGCGATGATATCTTTGGTGAATCGAGAGGTAAAGT 720
 QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgLysCys 260
 Db 721 TACCATCAAGAAATTTTGACCTGTGTGAGTTTGAGAGCGCGGAGAACTCCGCAATGC 780
 QY 261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
 Db 781 AAGGAGGCTTACAGGGAGGAGTTTTCATATATCTTGAAATCAAAATTAATGATCTCAAC 840
 QY 281 AspGluLeuProProGluGluAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
 Db 841 GATAGTTCCTCTCTGAGCAAGACATATAGTTTATCTGATGAAACGGTCAAG 900
 QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
 Db 901 ACAATATGTTCAATATATGATCTTCAATGTGATCTTCTCTCAATCTTACTGGAATTCCT 960
 QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
 Db 961 GGGTCTATGAGGCCCTCAAGAGGTACGAATGAGCGTTCAATATATGATGCAATCAAT 1020
 QY 341 AspValValGlyLysGluProAsnThrPargArgProSerAlaAsnGlyThrGlyIleLeu 360
 Db 1021 GACGTTGTGTAAAGAGCCAAACTGAGAGAGGCCATCTGCAATATGAACTGGAAATATCTC 1080
 QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
 Db 1081 CCACTCTCAACAGGAGACTGGAATCTTCCATCTCTGCAACAGGTACGAGGCGCCCTTATG 1140
 QY 381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgIleSerLys 400
 Db 1141 AGGTCAAACCCCATCTTGGGAAGATATGATCTGCTCAGTTGGTTCAGAGAAGTAA 1200
 QY 401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
 Db 1201 CGAACAGAGAGATGATGATGATGAGCGGTTTACTTCCAAATGGGTTGTCATATAGGTTA 1260
 QY 421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
 Db 1261 TTTGGTCAAGGTTCAATGCTGCTAGACCTATCCGAGGCCCTCCGATTCATCATCAATA 1320
 QY 441 TrpArgIleMetIleAlaValysGlyIleThrProValCysCysAlaArgCysValProMet 460
 Db 1321 TGGAGAGATGATGATGATGATGAGGTTGAACTCCGCTGTGTGCTGTGTATCTTATG 1380
 QY 461 GlyLysGlyIleGluThrLysLeuProGluValAlaAsnCysSerAlaArgThrAspLeu 480
 Db 1381 GGAAAGGAGGATGAAACTAACTGAGTGCATGATGTTTCAAGCAAGAACTATTTG 1440
 QY 481 AsnMetLeuAlaValHisIleTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500

```

Db      1441 AATATCTGCTTAACATTACGCCGTTGCATTGGATGTGAGATCTGTTTTTTCGTAACA 1500
Qy      501 ACPATGUGLUGAAspPhealaserTyrThrGluPheLeuArgTyrLeuSerSerylasp 520
Db      1501 GACAGGGAAGAAGATTTTGGCTTACACTGAATTTCTCCGGATACCTTAGCTCAAAAGAT 1560
Qy      521 ArgAlaGlyValAlaIleValLeuAspArgIleThrThrLeuPheLeuValProProSerArg 540
Db      1561 CGGGGGGGGTGGCCAAATTAGATGATGACAACTTATCTTGGTGGCTCCATCAGAT 1620
Qy      541 PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys 560
Db      1621 TTCTTAACCTGATTAATCCAGAGACCCGTCAGAAACGGCTAATAGTGTGTTCTCAAG 1680
Qy      561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGlnSerGlnSerAsnPro 580
Db      1681 TTACCCCGCCAGCCGTTCCGTGTTACAGATCATACAGACAGAAATCTCAGTCAATCCT 1740
Qy      581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
Db      1741 CTGCATTAAATGATGATAGAGCCCGGATTCACCTGCCAATGCTAGTACAGTTATATCCT 1800
Qy      601 ProArgGluAsnTyrIleArgGlyAlaProGluHisIleuThrAlaAlaSerLysProSer 620
Db      1801 CTTAGGGAATAATTAACATTAGGGGTGACACAGAACATTTGACAGCTGCTTCAAAACCATCT 1860
Qy      621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
Db      1861 GTTAGCGAGCTCTCAGAAATACCTTAATATGACGCCCTCAACCTGGGGTTAGTTAACT 1920
Qy      641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
Db      1921 CCGAGCTTTTACCCACTGCGATCTATTTCTCCGCAACTTCTAACCTGTCGCCCT 1980
Qy      661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
Db      1981 GAGAGTCACCAACCTATGTGACGACCTTCAACAGTTGTTTCCACAGACATCATGTCAT 2040
Qy      681 GlyLeuTyrAsnGlyAlaAlaProSerGlnAlaIleTyrLysArgGlyProGlnThrValHis 700
Db      2041 GGCCTGTACATGAGAGAGACCGCTCTCAAGCTTGGAAAAAGAGTCCACAAACAGTTCAAT 2100
Qy      701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
Db      2101 GATGCTCAATATGATCATTTCCAAACATACGAAATCAGATACCTCAGCTGGCAACATA 2160
Qy      721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
Db      2161 CTCTCTCTCTCTCGCGTTACCTTCAAGCTTCAAAACACCCCACTACCTAGTGGAAATG 2220
Qy      741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
Db      2221 GTTCATATGGCAATGCAATACAGAGCAATCTGTTAAATGCTCAGGCTGCTCCGTTA 2280
Qy      761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
Db      2281 CCAAAATATGCTCATATATATTAATTCATCATGTACACAGGGTTCGCAATCATCTCTGTT 2340
Qy      781 SerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGly 800
Db      2341 TCTCAGCCCATGTCCAGCAATACCAACAGAGCGCTCCATGCCAAACCAAACTATGCT 2400
Qy      801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
Db      2401 CCAATTCCAAGTATATGCAAGCTAATTTTCATGGGTAAACAACAATCAGGACAGAAC 2460
Qy      821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaLeuLeuGlu 840
Db      2461 TTAAACCTTCCCAATTTCAAGTGCATGCAACCAACAGCAATATAGGCAATTTTAGAG 2520
Qy      841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860

```

```

Db      2521 CCACAAACCAAGACATACGATTGCGACCTATGATCTCGGGGATGTGTCAGGGTACAACA 2580
Qy      861 AspGlyGlyValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
Db      2581 GATGGGGAGGTGATAGATACAGATACCAAGTCAACATCAATTTGACAGCAACCTT 2640
Qy      881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
Db      2641 CTCTCCAGATACAGAGAACAGACAGCAAGCTTTCAGTACTCCGGCTGACAGGGG 2700
Qy      901 Pro 901
Db      2701 CCT 2703

RESULT 2
ADN73104
ID      ADN73104 standard; cDNA, 3171 BP.
XX
AC      ADN73104;
XX
DT      15-JUL-2004 (first entry)
XX
DE      Thale cress cDNA upregulated in E2Fa/Dpa expressing plants SeqID 999.
XX
KW      gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;
KW      growth regulator; animal feed product; thale cress;
KW      cell wall biosynthesis; nitrogen metabolism; carbon metabolism.
XX
OS      Arabidopsis thaliana.
XX
PN      MO2004035798-A2.
XX
PD      29-APR-2004.
XX
PF      20-OCT-2003; 2003WO-EP011658.
XX
PR      18-OCT-2002; 2002BP-00079408.
XX
PA      (CROP-) CROPDESIGN NV.
XX
PI      Inze D, De Veylder L, Vlieghe K;
XX
DR      WPI: 2004-348466/32.
XX
PT      P-PSDB; ADN73105.
XX
PT      Altering plant characteristics, useful for producing plants for enzyme or
PT      pharmaceutical production comprises modifying in a plant, expression of
PT      one or more nucleic acids and/or modifying level or activity of one or
PT      more proteins.
XX
PS      Claim 1; SEQ ID NO 999; 134bp; English.
XX
CC      This invention relates to a novel method for altering one or more plant
CC      characteristics. Specifically, it refers to identifying genes that are up
CC      - or down-regulated in transgenic plants overexpressing the heterodimeric
CC      E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to
CC      alter plant characteristics accordingly. The present invention describes
CC      generating transgenic plants for the production of growth regulators,
CC      enzymes, therapeutics, pharmaceuticals and animal feed products, where
CC      the altered plant characteristics are selected from increased yield or
CC      biomass, enhanced survival capacity, stress tolerance, plant architecture
CC      or physiology, altered endoreduplication, biochemistry, signal
CC      transduction, storage lipid mobilisation and/or altered photosynthesis,
CC      each relative to the corresponding wild type plants. Accordingly, these
CC      sequences can also be useful as positive or negative selectable markers
CC      during transformation of cells or tissues. The identified genes play a
CC      role in a variety of biological processes such as DNA replication, cell
CC      wall biosynthesis, nitrogen and/or carbon metabolism or they function as
CC      transcription factors. This polynucleotide sequence is thale cress cDNA
CC      upregulated 1.3 fold or more in plants overexpressing the E2Fa/Dpa
CC      transcription factor, given in an exemplification of the invention.
XX
SQ      Sequence 3171 BP; 910 A; 723 C; 709 G; 829 T; 0 U; 0 Other;

```

Alignment Scores:

Pred. No.:	2,13e-238	Length:	3171
Score:	4656.50	Matches:	892
Percent Similarity:	85.04%	Conservative:	6
Best Local Similarity:	84.47%	Mismatches:	3
Query Match:	96.99%	Indels:	155
DB:	12	Gaps:	3

US-09-920-705-3 (1-901) x ADN73104 (1-3171)

```
QY      1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
DB      1 ATGCCGTTATCTATGAAAGCCATTCAGAGCCAGATGATCCGGTTTCAGTCAAACTCTT 60
QY      21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
DB      61 TGGGTCGGTACGCTTAACGCCGAGACGACAGCTCAGATCTGACCGAGTTGTTGGAGA 120
QY      41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyr 60
DB      121 TACGGCATATGTATGATACACGGTGTATTTCTTCAAGAGCTTTCGCTTATATATCTAC 180
QY      61 ArgHisValGluGluValAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
DB      181 AGACATGTGAGAGAGACAGTCGACGCCAAGAGGCTCTTCAAGAGCAAAATTTGAATGGA 240
QY      81 Ser----- 81
DB      241 ACGAATGAGCTCTGCAAAAGAAATACATCAAAACAGTTTACTCTCGTTTATGCCAGAA 300
QY      81 ----- 81
DB      301 GATTTAATCCGATTGCTTATCTGCAAACTTGTAATCCGCTTTAGCAGTCTGATTGCA 360
QY      81 ----- 81
DB      361 GGTTCCTTCAGATGTTAAATCTCACAGCTTGAGTGAATGATATGCGTACTCCACTC 420
QY      82 -----GlnIleLysIleGluTyr----- 87
DB      421 GAAATTCGCAATGATAAAGAAATTTTGGAGTTTAAAGTTCAAGTTTCTCTGCAAAATCAAA 480
QY      87 ----- 87
DB      481 CCGCGTGAAGAAATTTGTCTATGCGGTACGTAGATATCATATACATGATCGCAACA 540
QY      87 ----- 87
DB      541 GAAATTTCTGGCAACCGTATCGTCTAATCTAGTATCGTGGACATATTCATATCAG 600
QY      87 ----- 87
DB      601 TGTGAGAGAGTTTGGGCTTGTTTTGGCTTCGATGCTTCGAGAAAGCAAAATTAAGTGG 660
QY      87 ----- 87
DB      661 TTAAGCGGCTAATCTTACCAATACGAAACGCTTATTCGAAAGCAACATGTTTGTCTTC 720
QY      88 ---AlaArgPro---AlaLysProCysLysSerLeuThrValGlyValIleGlyProAsn 105
DB      721 AATTCCTCAATCCCTAGCGCAAAACCTTGTAAAGATCTATGGGAGTGGGAAATCGGCTTAT 780
QY      106 ValSerLysAspAspLeuGluGluLysPheSerLysPheGlyLysIleGluAspPheArg 125
DB      781 GTCTCCAGAGATGACCTGGAGAGAGAGTTGACAGAGTTTGGGAAATAGAGGATTTTACG 840
QY      126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
DB      841 TTTCCTAGAGAAAGCAAGACAGCTTTCATATGATTTATGAGATGATGATGATGCTTTACAG 900
QY      146 AlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArg 165
```

```
DB      901 GCTAAGACATGATGAAAGCCCTATGGGTGTAACCTTTTGGCTTGTATTTCTCCGG 960
QY      166 SerGlnAlaProLysLysGluGlnTPrAlaGlySerTyrAspAsnArgAsnGlyAsnMet 185
DB      961 TCACAAAGCCCAAAAAAGAACATGGCTGGCTCTTACGATTAACAGAAATGGCAATATG 1020
QY      186 AsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGlnProSer 205
DB      1021 AATCATTAACCGCAGATATCTCATCATATGAAAGCTTTAAAGAGATGTCAGCAAGT 1080
QY      206 LysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeu 225
DB      1081 AAGGTTCTGTGGATTTGGGTTCCCTCTCTCTCTCATCATCATCATGATGAGCAAAATTTG 1140
QY      226 HisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsn 245
DB      1141 CACAATGGCAGATGATCTCTTGGTGAAGATCGAGAGGGTAAAGATTAACCATCAAGAAAT 1200
QY      246 PheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGln 265
DB      1201 TTTGCACTTTGTGAGATTGAGAGCGGAGAGAGCTGCCAATGCAGAAAGGCTTACG 1260
QY      266 GlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProPro 285
DB      1261 GCGAGGTTATTCAATTAATCTTAATCAAAATTAATGATCTAAACGATGAGTTCCTCT 1320
QY      286 GluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsn 305
DB      1321 GAGCAAGACGATACGATGTTTACTCTGTATGAAAGCGTCAGAACGACATATTTCAAT 1380
QY      306 AsnAspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgPro 325
DB      1381 AATGATCTCTATTTGTAATCTTCTCTCTCATTTCTGGAATTTCTGGGCTCTAGAGCC 1440
QY      326 LeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValAlaGlyLys 345
DB      1441 CTCAGAGGTACAAATGAGCGTTTCATATATGTGTCAAAATACATGACGTTGTGGTAAG 1500
QY      346 GluProAsnTrpArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGly 365
DB      1501 GAGCCAAACGTGAGAGGCGCATCTGCAAAATGGAATCGGAATCTCCATCTCCAAACGGA 1560
QY      366 ProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAsp 385
DB      1561 CTTGGAATCTCCCATCATCTCTGCAACAGTACAGAGGCGCTTATGAGGTCAAAACCCGAT 1620
QY      386 SerTrpGluGlyTyrAspProAlaGlnLeuValArgGlnSerLysArgThrArgArgAsp 405
DB      1621 TCTTGGAGAGATATGATCTCTGCTCACTGCTGTCAGAAAGTAAACGAACCAAGAGAT 1680
QY      406 GlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
DB      1681 GGATCAAGTGAAGGTTTACTTCCATAGGTGTGTGATAGAGGTCATTTGTGTCAGGTTCA 1740
QY      426 ValAlaAlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIle 445
DB      1741 GTTGCTGCTAGACCTATCCGTGGGCCCTCGATTCGATACATATGAGAGAAAGATT 1800
QY      446 AlaLysGlyGlyThrProValCysGlyAlaArgCysValProMetGlyLysGlyIleGlu 465
DB      1801 GCCAAGGGTGAACCTCCGCTCTGTGTGCTCGTTGTGTACTATGCGAAAGGGGATTTGAA 1860
QY      466 ThrLysLeuProGluValValaAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
DB      1861 ACTTAATGCTGCTAGAGTTCGCAATTTGTTTCGCAAGAACTGATTTGAATATGCTCGCTAAA 1920
QY      486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
DB      1921 CATTAAGCCGCTGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
QY      506 PheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAla 525
DB      1981 TTTGCGTTTAACTGAATTTCTCGGTACCTTAAGTCAAAAGATCGGCGGTGTTGGCC 2040
```

QY 526 LyeLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspVal 545
 DB 2041 AATTTAGATGATGATGAACCTTATCTTGGTCTCCATCAAGATTTCTTAACCTGATGA 2100
 QY 546 LeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuIleuPheProProPheVal 565
 DB 2101 CTCGAGTGAACCGCTCAAGAACCGCTATATGCTGTCTTCAAGTTAACCCCGCCAGCC 2160
 QY 566 ValProValThrAlaSerTyrArgGlnGlnSerGlnSerAsnProLeuHisTyrMetAsp 585
 DB 2161 GTTCTGTTACAGCATCAATCAAGCAAGATCTCAGTCCATCTCTGCAATTAATGAT 2220
 QY 586 GlnAlaArgSerProAlaAsnAlaSerHisSerLeuTyrProProAlaGluAsnTyr 605
 DB 2221 CAAGCCCGGATTCACCTGCAATGATGATCAAGATTAATCTCTGAGGAAATATAC 2280
 QY 606 IleArgGlyValAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluProLeu 625
 DB 2281 ATTAGGGGTGCACAGAACATTTGACAGCTGCTTCAAAACCATCTGTAGCGAGCTCTC 2340
 QY 626 ArgIleProAsnAlaAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuAla 645
 DB 2341 AGAATACCTAATATGACAGCGCTCAAGCTGGGGTTAGTTAACTCGAGCTTTAGCC 2400
 QY 646 ThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnPro 665
 DB 2401 ACTCTGGCATCTATCTCCCTGCAACTCTCAACCTGCTGCCCTGAGATCCCAACCT 2460
 QY 666 MetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGly 685
 DB 2461 ATGTCAAGACCTTCACAGCTGTTTCCACAGCATCACTGCAATGATGATGATGATGATG 2520
 QY 686 GlnAlaProSerGlnAlaTyrIleArgGlyProGlnThrValHisAspAlaSerAsnGln 705
 DB 2521 GAAGCAACCGCTCAAGCTTGAAAGAGGCTCCCAAAACATTCATGATGATGATGATGATG 2580
 QY 706 SerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProSer 725
 DB 2581 TCATTCACAAATACGAAATACGATCACTCCAGCTGGGGACTACTCTCTCTCTCTCG 2640
 QY 726 ArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMet 745
 DB 2641 CGTACCTCCACCTTCAACCAACCCCACTACATGATGATGATGATGATGATGATGATG 2700
 QY 746 GlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHis 765
 DB 2701 CAATACCAAGCCCAATCTGTTAATGCTCAGCTGCTCGTTACCAATATGCTCAT 2760
 QY 766 AsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetVal 785
 DB 2761 AATTAATTAATTCATGTAACCTCGGGTTCCTCAAAATCATCTTTCTCGAGCCCATGCTC 2820
 QY 786 GlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyr 805
 DB 2821 CAGCATACCAACCAAGGCTCAATGCCCAACCAAACTATGCTCAATTTCCAAAGTTAT 2880
 QY 806 GlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGln 825
 DB 2881 CAGCAAGCTAATTTTTCATGCGTAAACAAATCAAGCAAGAACTTAAACCTTCCCA 2940
 QY 826 PheGlnAlaAlaMetGlnProProAlaAspValAlaAsnLeuGluProGlnAsnGlnAla 845
 DB 2941 TTTCAAGCTGCTCAATGACCAACCAAGATTAAGCAAAATTTTAAGCCCAAAACCAAGCA 3000
 QY 846 LeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAsp 865
 DB 3001 CTACGATTCAGCTATGATCTCTGGGATGCTCAGGGTCAACACATGGGAGGCTCAT 3060
 QY 866 LysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGln 885
 DB 3061 AAGAAATCAGATACCAAGTCAACATCAATTTGACAGCAAACTCTCTTCCAGATACAG 3120

QY 886 GlnTyrGlnGlnGlnInserSerGlyThrProAlaGlyGlnGlyPro 901
 DB 3121 CAGAAACAGACAGACAGCTTTCAGGTACTCCGGCTGACAGGGGCT 3168
 RESULT 3
 ID ABA95166 standard; DNA; 4593 BP.
 AC ABA95166;
 XX 20-MAY-2002 (first entry)
 DT
 DE Arabidopsis floral induction gene (FPA) sequence.
 XX FPA, floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KM flowering locus C; transgenic; gene; ds.
 XX Arabidopsis thaliana.
 OS
 PN MO200212518-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 02-AUG-2001; 2001WO-US024427.
 XX
 PR 03-ANG-2000; 2000US-0222550P.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Amsino RM, Schomburg FM, Michaels SD, Patton D;
 DR WPI; 2002-227160/28.
 DR P-PSDB; ABB07659.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 PT
 PS
 XX Claim 5; Page 27-29; 39pp; English.
 XX
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, frigida (FRI) and FLC which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FPA gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene sequence
 XX
 SQ Sequence 4593 BP; 1275 A; 958 C; 978 G; 1382 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3, 64e-224 Length: 4593
 Score: 4392.00 Matches: 899
 Percent Similarity: 58.72% Conservative: 0
 Best Local Similarity: 58.72% Mismatches: 2
 Query Match: 91.48% Indels: 631
 DB: 6 Gaps: 4
 US-09-920-705-3 (1-901) x ABA95166 (1-4593)
 QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 DB 1 ATGGCGTTATCTATGAGCATTCAGAGCCGATGATTCGGGTTCCAGTCAAAACATCTT 60
 QY 21 TrpValGlySerLeuThrProGluThrThrGlnSerAspLeuThrGluLeuPheGlyArg 40

Db	61	TGGGTGGGTAGCCTTAACGCCGAGACGACAGACTCAGATCTGACCGAGTTGTTGGAAGA	120
QY	41	TYRGLYASPILEASPARGILETHVALYRSESRARGLYPHEALPHEILETYR	60
Db	121	TACGGCGATATGTATGAATCAGCGGTATCTTTCACGAGCGTTTGCGTTATATACTAC	180
QY	61	ARXHLVGLGLUHLUHLAVALAIALVYGLUHLALEUHLNGLYALASHLVLSHLGLY	80
Db	181	AGCAATGTGAGGAGACGACGTGGACCAAGAGAGCGCTCTTCAGAGCCAAATTTGAATTGA	240
QY	81	SESGNILEYSELGLUYRALARGPRO-----	90
Db	241	AGTCAATTAAGATCGAATACGACACGCGTTTGTTCTTATATATCTTCGTTGTTTC	300
QY	90	-----	90
Db	301	TCCTAACTTGATTCCTTTTGTTCACAGATTATCTCTTTTGGCAATTCATAGTCCAGGT	360
QY	90	-----	90
Db	361	TCACAAACTTGTATGATGCTGTGTTAATGCCAAAAATTCCTGTGTAATCTGTTTTTTT	420
QY	90	-----	90
Db	421	TTCCATCGTACAAATCAAAAGTGGAACTGTTTTTTTCTATATATAGCTGTAGCTT	480
QY	90	-----	90
Db	481	AAGGCGAAACCTGATCCGATCGAAGACGTCTTTCTCAAAATTACTTGGTTAATCGACT	540
QY	90	-----	90
Db	541	CGCGCAAGCCAAACACAGAGAGAGCTCTGCAAAATTTGATGTTAAGCATATATACTTC	600
QY	90	-----	90
Db	601	TTAGCAATGAGCTCTGCAAAAAGATATACATCAACACATGTTTACTCTCGTTTATGCGAA	660
QY	90	-----	90
Db	661	GAAAGTTTAATCCGATTCGTGTTTATCTGCAAACTGTATCCGCGTTTAGCAGTCGATT	720
QY	90	-----	90
Db	721	TCAGGTGCGCTTCAGATGTTAATCTCACAGCTGAGTATGAATGTATTGGCTACTCCA	780
QY	90	-----	90
Db	781	CTCGAAATTCGCAATGCTGMAAATTTGGAAGGAGCAACCTGTATCATGCGCAACACAGA	840
QY	90	-----	90
Db	841	TAAGAAATTTGAGTTTAAAAAGTTCAGTTTCTGCAAAAATCAACCGCGTGAGAAAT	900
QY	90	-----	90
Db	901	TTGTCTATGGCGGTACGTATATCATATACATCTGCAATTCGCAACAGAAAAATTCGGCAA	960
QY	90	-----	90
Db	961	CCGATTCGTCTTAATCTCTAGTATCGCTGGCACATATCATATCAGTGTGAGGAGTTT	1020
QY	90	-----	90
Db	1021	GGGCTTGTTTGGCTCGATGCTTCAGGAAGACAAATTAAGTGTTGTAAGGCGGCTAAC	1080
QY	90	-----	90
Db	1081	TCTACCAATCAGAAAGCGTTTATTTGGAAGAACCATGTTTGTCTCAATTCGCATCCCTA	1140
QY	90	-----	90

Db	1141	CGTCAATCTGGGCTTTCCCATTTGATGTTCTTAGGAAGTAGACTCTTCAACAACATTC	1200
QY	90	-----	90
Db	1201	TGGGATTGGTACAGTTGCGAGTGGACACAAATATTCATTTCAGGCGAGACTCAATTT	1260
QY	90	-----	90
Db	1261	CAATGCTTGGCCGCTTTGAGTACTTAATGAGAGTTTCAGGAGTTGGTTATTGGGAC	1320
QY	90	-----	90
Db	1321	TCAGGCAATGGATTAAGACAGGAATGTTTTCAGAAAGTATTAATGTCCTTCCGGCTTGG	1380
QY	90	-----	90
Db	1381	GGGACTACATCTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAAACTATTACA	1440
QY	90	-----	90
Db	1441	CTGATTTAATAGAGAAAGAGAGATGTTTCTCTGGCGCAATTTATTTGTTAGGAA	1500
QY	90	-----	90
Db	1501	AGCAATTGATATGAATTGGTGTGCTAGTGTATTAAGTAATTACTAGTAGTTTGTGTGT	1560
QY	90	-----	90
Db	1561	AGTTTCCTGATGTTTGAATGCTTATTCCTGGCAACCTATCTGGGTAGTATCGCTCTT	1620
QY	91	-----	103
Db	1621	ATCGACCTTTTCTTGTTCAGGCGCAAAACCTGTATGAAGTCTATGGGTGGGGAATCGGC	1680
QY	104	ProAenValSerIysAspAspLeuGluGlnPheSerIysPheGlyIysIleGluAsp	123
Db	1681	CTTATGCTTCACAGAGATGACCTGGAGAGAGATTCAGCAAGTTTGGGAAATTCGAGGAT	1740
QY	124	PheArgPheLeuArgGluArgIysThrAlaPheIleAspIyrTrjGluMetAspAspAla	143
Db	1741	TTTAGGTTTTCACAGAAACCGAAGACAGCTTTCATTATTTATTTATGAGATGATGCT	1800
QY	144	LeuGlnAlaIysSerMetAengIyLysProMetGlyIysSerPheLeuArgValAspPhe	163
Db	1801	TTACAGGCTAAGACATGAAATGAAACCTATGGGTGTAGCTTTTGGCGTGTGATTTT	1860
QY	164	LeuArgSerGlnAlaProIyLys-----	171
Db	1861	CTCGGTCACAGCGCCAAAAAAGTAAAGACCTCTTGCGCATTGATTTTACTTTTGA	1920
QY	171	-----	171
Db	1921	AAAGCTCCAGTAACAATTTGTTTATGTTCAATATTTGGCTCAAACTGATAGGGCTGAG	1980
QY	171	-----	171
Db	1981	CTGTCTGTGGCCCTAGAGCAGATTTACTGTCTATTTTCATTTAGAGTAGGCT	2040
QY	172	-----	176
Db	2041	CAACTCTTAAGTCGTGAATCAAGTTACCTTTGTGTTATCTTCAGGAACAATGGGCTGGC	2100
QY	177	SerTrjAspAsnArgAengIyAsnMetAsnH1LysProGln-----	190
Db	2101	TCTTACGATACAGAAATGGCAATATGAATCAATAAACCGA-GTTATGCTTGAATGTTGG	2159
QY	190	-----	190
Db	2160	AAAGTATGCTCTTGTTACTAGTATGATATGATAGTTACTGTTTTTGAAGTTTTGTTATTA	2219
QY	191	-----	207
Db	2220	TTCTTACAGTATCTCTCATATATGAAGACTTTTAAAGGAGATGTCAGCGCACTAAGGTT	2279

QY 208 LeuTrpIleGlyPheProThrAlaThrGlnCysAsnAspGlnIleLeuHisAsn 227
DB 2280 CTGGGATGAGTGGTCCCTCCTACCTGTAACAATGCAATGAGCAAAATTCGACAAT 2339
QY 228 AlaMetIleLeuPheGlyGlnIleGluArgValIleSerTyProSerArgAsnPheAla 247
DB 2340 GCGATGATACCTCTTGGTGAAGTGAAGAGGTAAAGTTACCACTCAAGAAATTTGCA 2399
QY 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysIleGlnIleGluGlnIleArg 267
DB 2400 CTGTGAGATTGAGAGCGGAGGAGAGCTCGCAATGCAAGGAAAGGCTTACAGGGAGAG 2459
QY 268 LeuPheAsnAsnProArgIleValIleMetIleSerAsnAspGlnLeuProPheGln 287
DB 2460 TTATTCATATATCTAGAAATCAAAATTAATGACTCAAAACATGAGTTCCTCTGAGCAA 2519
QY 288 AspAspThrSerPheTySerGlyMetIleArgSerArgThrAspMetPheAsnAsnAsp 307
DB 2520 GACGATACGATGTTTCTCTGGAATGAACGGTCAGACAGACATATGTTCAATATGAT 2579
QY 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
DB 2580 CCTTCATTTGATCTTCTCTCCTCATTTCTAGCAATTCCTGGTCTATGAGGCCCTCAGA 2639
QY 328 GlyThrAsnGluArgSerTyArgGlnValGluThrAsnAspValAlaGlyIleArgPro 347
DB 2640 GGTACAAATAGAGGCTTATATATGATGAGCAAAATACATACGTTGTGTAAAGAACCA 2699
QY 348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
DB 2700 AACTGAGAGAGGCAATCTGCAATGAACTGAAATACCTCCATCTCCAAACGACCTCGA 2759
QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
DB 2760 ATCTCCCATCTCTCGACAAAGGATACGAGCGCCCTATGAGTCAAAACCCGATTTCTGG 2819
QY 388 GlyGlyThrAspProAlaGlnLeuValArgGluSerTyArgTrpArgAlaGluArgPheGlySer 407
DB 2820 GAAGGATATATCCTGCTCAGTGTGTGAGAAAGTAAACGAAACCAAGATGATGATCA 2879
QY 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
DB 2880 GTGACGCTTTTACCTCAATGCGGTGATGAGAGGTCATTTGGTCAGAGTTCAGTTGCT 2939
QY 428 AlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaIle 447
DB 2940 GCTAGACCTATCCGTGGCCCCCTGATTCGATCATATGAGAGGAATGATGCAAG 2999
QY 448 GlyGlyThrProValCysGlyAlaArgCysValProMetGlyIleGluThrIle 467
DB 3000 GGTGGAATCCCGCTGTGTGTGCTGTGTGTACTATGGGAAAGGGGATTGAAACCTAA 3059
QY 468 Leu----- 468
DB 3060 CTGTGAGTACTAATTTTACGACTTTAACCTCTTACTGTTTCTTTTACAGACCATTT 3119
QY 468 ----- 468
DB 3120 ATATATTTTCATTTTCATTCGATGAGAGTAACATTAATAGATAGTACATTTTATTT 3179
QY 468 ----- 468
DB 3180 TTACTATTACTGTTAGTTTCTGAGATGCTGATTTTTCATGCTGTGATGATTTTGG 3239
QY 469 -----ProGluVal 471
DB 3240 GCATGGCCCTCAATACTGACTTTGTTTTTTAAATATGATTTATAGCCCTGAGGT 3299
QY 471 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaIleHisIleTyAlaValAlaIle 491
DB 3300 CGTCAATTTGTTAGCAAGAACTGATTTGAATATGCTCGTAAACATTAACCCCTGTGCAT 3359

QY 491 eGlyCysGlnIleValPhePheValProAspArgGlnIleAspPheAlaSerTyThrGln 511
DB 3360 TGGATGTGAGATGCTGTTTTCCTACCAACAGGAAAGAAATTTGGCTTTACCTGA 3419
QY 511 uPheLeuArgTyLeuSerSerTyAspArgAlaGlyValAlaIleAspAspGlyTyThr 531
DB 3420 ATTTCTCCGGTACCTTAGCTCAAAAGATCGGGGGGTGTGTGCAAAATTTGATGATGCTAC 3479
QY 531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGln 551
DB 3480 AACTTATTTCTGGTGCCTTCATCAGATTTCTTAATCTGATGATCACTCAAGTGCCTGCTCA 3539
QY 551 nGluArgLeuTyArgValValLeuIleuIleuProProAlaValProValThrAlaSer 571
DB 3540 AGAACGCTATATGATGTTGTTGTTCTCAAGTTTACCCCGCAAGCCGTTCTCTGTACAGCATC 3599
QY 571 rTyArgGlnGluSerGlnSerAsnProLeuHisTyMetAspGlnAlaArgAspSerPr 591
DB 3600 ATACAGACAAGATCTCAATCCATCTCTGCAATTATGATCAAGCCCGGATTCACCC 3659
QY 591 oAlaAsnAlaSerHisSerLeuTyProProArgGluAsnTyIleArgGlyValAlaProGln 611
DB 3660 TGCCAAATGCTAGTCAACAGTTTATCTCTCTGAGGAAATTCATTAGGGGTGACACAG 3719
QY 611 uHisLeuThrAlaAlaSerTyProSerValSerGluProLeuArgIleProAsnAsnAl 631
DB 3720 ACATTTGACAGCTGCTTCAAAACATCTGTATAGCGACCTCTCAGAAATCTTATATATGC 3779
QY 631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuAlaThrLeuAlaSerIleLe 651
DB 3780 AGCGCTCAAGAGTGGGTTATGTTTATCTCGAGGCTTTTAGCACCTGGAATCTATTTCT 3839
QY 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerThr 671
DB 3840 CCTCGCAACTTCTCAACCTGCTGCCCTGAGAGTCAACCACTATGTGAGACCTTCAAC 3899
QY 671 rValAlaSerThrAlaHisGlnSerAsnGlyLeuTyArgGlnGlyValAlaProSerGlnAl 691
DB 3900 AGTTGTTTCCACGACATCATGTCATGACATGACATGACATGAGAAAGCAACCGTCTCAGC 3959
QY 691 aTrpIleArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyArg 711
DB 3960 TTGGAAGAAAGGTCACAAACAGTTCAATGATGCTCAATTCAGTCTTCCAAACATACCG 4019
QY 711 yAsnGlnTyThrProAlaGlyGlnLeuProProProSerArgTyProProAlaSer 731
DB 4020 AATTCAGTACACTCCAGCTGGGCACTACTCTCTCTCTGCGTTACCTCCACTTC 4079
QY 731 rAsnAsnProAsnTyThrSerGlyMetValHisGlyAsnMetGlnTyGlnSerGlnSe 751
DB 4080 AAACAAACCCCACTACATGAGTAAGTAAGTCCATGCAATGCAATACCAAGCCCAATC 4139
QY 751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTySerMetTy 771
DB 4140 TGTTAACATGCTCAGCTGCTCTCTCCGTACCAATATGCTCATTAATATTTCCATGTA 4199
QY 771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyGlnProGln 791
DB 4200 CACTCAGGGTTCGTCAATCATCTGTGTTCTGAGCCCACTGTCAGCAATACCAACCA 4259
QY 791 uAlaSerMetProAsnGlnAsnTyArgIleProIleProSerTyArgGlnAlaAsnPheHis 811
DB 4260 AGGCTCATGCAACCAAACTATGATGCTCAATTCCAATGTATACGAAAGCTTATTTTCA 4319
QY 811 eGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGln 831
DB 4320 TGGCGTAACAACAATCAGGACAGAACTTAAACCTTCCAAATTTCAAGCTGCCATGCA 4379
QY 831 nProProAlaAspGlyValAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
DB 4380 ACCACACGAGATTAAGCAAAATTTAGAGCCCAAAACCAAGCACTACGATTCAGACCTAT 4439
QY 851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspIleAsnGlnArgTyArg 871

Db 4440 GATCTCGGGATGTCAGGGTACACAGATGGGAGTGCATAGATCAGATACCA 4499
 QY 871 nserThleuGlnpheleAlaaleuLeuLeuGlnGlnGlnGlnGlnGlnGln 891
 Db 4500 GTCAACACTACAAATTTGCAGCAACCTCTTCTCCAGATACAGCAAGAACGACAA 4559
 QY 891 nserSerGlyThrProAlaGlyGlnGlyPro 901
 Db 4560 GTCTTCAGTACTCCGCTGACACAGGGCCT 4590
 RESULT 4
 ID ABA95168/c
 AC ABA95168 standard; DNA; 801 BP.
 XX ABA95168;
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) antisense fragment.
 XX
 KM FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KW flowering locus C; transgenic; gene; antisense; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200212518-A2.
 PD 14-FEB-2002.
 PF 02-AUG-2001; 2001MO-US024427.
 XX
 PR 03-AUG-2000; 2000US-0222550P.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
 DR WPI; 2002-227160/28.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 XX
 PS Claim 5; Page 37; 39pp; English.
 XX
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, frigida (FRI) and FLC which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene antisense fragment
 XX
 SQ Sequence 801 BP; 206 A; 207 C; 166 G; 222 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.51e-68 Length: 801
 Score: 1436.00 Matches: 265
 Percent Similarity: 99.62% Conservative: 0
 Best Local Similarity: 99.62% Mismatches: 1
 Query Match: 29.91% Indels: 0
 DB: 6 Gaps: 0

US-09-920-705-3 (1-901) x ABA95168 (1-801)

QY 197 AappheleVgLyAspValGlnProSerLybValLeuTrpLleGlyPheProProThra 216
 Db 799 GACITTTAAAGAGAGATGTCAGCCAAAGTAAAGTTCTGTGATGGGTCTCTCTACTGCT 740
 QY 217 ThrGlnCyAspAspGlnGlnLleuLhisAsnAlaMetLleuPheGlyGlnLleu 236
 Db 739 AONCAATGCAATGATGAGCAATTTTGCACAAATGCCAATGATGATCTTTTGATGATGAG 680
 QY 237 ArgValLysSerTyrProSerArgAsnAspHeaLaleuValGlnPheArgSerAlaGln 256
 Db 679 AGGTAAATAAGTTAACCTCAAGAAATTTTGACATTGTGAGTTAGAGACGCCGAGAA 620
 QY 257 AlaArgGlnCybVgGlnGlyLeuGlnGlnLysArgLeuPheAsnAsnProArgLle 276
 Db 619 GCTGCCCAATGCAAGAAAGGCCCTACAGGGAGGATTATCAATATCCATGAAATCAAAATT 560
 QY 277 MetTyrSerAsnAspGlnLeuProProGlnGlnAspAspThrSerPheTyrSerGlyMet 296
 Db 559 ATGTACTCAAGAGATGATGCTCTCTGACGAAACGATACTAGTTTACTCTGTGATG 500
 QY 297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCybValSerSerProHisSer 316
 Db 499 AAACGTCACAGGACAGATATGTTCAATATGATCTTCATTGTATCTTCTCTCATTTCT 440
 QY 317 ThrGlyLleProGlySerMetArgProLeuArgGlyThrAsnGlnuArgSerTyrAsnGly 336
 Db 439 ACTGSAATTCCTGGGTCTATAGAGGCCCTCAGAGGTACAAATVAGCGCTCATATATAGGT 380
 QY 337 AlaGlnTyrAsnAspValAlaGlyLysGlnProAsnTrpArgArgProSerAlaAsnGly 356
 Db 379 GCAGATACATCATGACGTTGTGTGTAAGAGCCAAACTGAGAGGACCATCTGCAATGGA 320
 QY 357 ThrGlyLleLeuProSerProThrGlyProGlyLleLeuProSerProAlaGlnGlyThr 376
 Db 319 ACTGGAATATCTCCATCTCCAAACAGACCTGGAATCTCTCCATCTCTGACAAAGTACG 260
 QY 377 ArgArgProMetCysSerAsnProAspSerTrpGlnGlyTyrAspProAlaGlnLeuVal 396
 Db 259 AGGCCCTCATGAGGTCAACCCGATTTCTTGGAGAGATATGATCTCTGCTCACTTGTC 200
 QY 397 ArgGlnSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyVal 416
 Db 199 AGAGAAAGTAAACCAACCAAGAGATGATCACTGACGGTTTACTCCAAATGGAGTGC 140
 QY 417 AspGlnArgSerPheGlyArgGlySerValAlaAlaArgProLleArgGlyProProAsp 436
 Db 139 GATGAGAGTCAATTTGTCAGAGTTGCTGCTGACCTTACCTGCGCCCTGAT 80
 QY 437 SerAspHisLleTrpArgGlyMetLleAlaLysGlyGlyThrProValCybValArg 456
 Db 79 TCTATATCAATATGAGAGAGATATTTGCCAAGGTGGAATCCCGCTCTGTGTGCTCCT 20
 QY 457 CysValProMetGlyLys 462
 Db 19 TGTGTACTTATGGGAAG 2
 RESULT 5
 ID ABA95170
 AC ABA95170 standard; DNA; 3715 BP.
 XX ABA95170;
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) fragment.
 XX
 KM FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KW flowering locus C; transgenic; gene; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200212518-A2.

XX 14-FEB-2002.
 PD 02-AUG-2001; 2001MO-US024427.
 XX 03-AUG-2000; 2000US-0222550P.
 PR
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 PA
 XX Amasino RM, Schomburg FM, Michaels SD, Patton D;
 PI WPI; 2002-227160/28.
 XX
 XX Novel isolated DNA sequence comprising coding sequence for floral
 PT induction gene (FPA) gene which controls flowering time in plants, useful
 PT for altering flowering time of plants and for downregulating flowering
 PT locus C (FLC) mRNA activity.
 XX
 PS Claim 5; Page 38; 39pp; English.
 XX
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, *frigida* (FRI) and *FLC* which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene fragment comprising
 CC the FPA promoter and intron
 CC
 SQ Sequence 3715 BP; 1055 A; 728 C; 711 G; 1221 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 1.97e-25 Length: 3715
 Score: 651.50 Matches: 171
 Percent Similarity: 27.23% Conservative: 0
 Best Local Similarity: 27.23% Mismatches: 0
 Query Match: 13.57% Indels: 457
 DB: 6 Gaps: 1
 US-09-920-705-3 (1-901) x ABA95170 (1-3715)
 QY 1 MetAlaLeuSerMetLysPProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 DB 1832 ATGGCGTTATCTATGAAAGCCATTCAGAGCGATGATTCGGTTCCAGTAAACAACTCT 1891
 QY 21 TTPValGlySerLeuThrProGluThrThrgLusSerAspLeuThrGluLeuPheGlyArg 40
 DB 1892 TGGGTGGGTAGCCTTAACGCCGAGAGACGACGAGTCGATCTGACCGAGTTGTTGAGAA 1951
 QY 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 DB 1952 TACGGCGATATGATGAAATCACGGTGTATCTTCAAGAGCTTTGGTTATATATCTAC 2011
 QY 61 ArgHisValGluGluValAlaValAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 DB 2012 AGACATGTGAGAGAGAGTGGCAGCCAAAGAGGCTCTTCAAGAGCAAAATTTGAATGA 2071
 QY 81 SerGlnIleLysIleGluTyrAlaArgPro----- 90
 DB 2072 AGTCAATTAAGATCGAATACGACACGCGTTTGTCTTATCTATATCTTCGTTGTTTC 2131
 QY 90 ----- 90
 DB 2132 TCTAATTGATGTCTTTTGTCAAGCATTAATCTCTTTTGGCAATTCATAGTCAGGT 2191
 QY 90 ----- 90

DB 2192 TCACAAACTTGTGATGATGCTGTTTATGTCAAAAATTTCTGTGATCTGTTTTTTT 2251
 QY 90 ----- 90
 DB 2252 TTCCTCATGTACAAATCAAGTCGAAACCTAGTTTTTTTCTATATATAGTCGTTAGCTT 2311
 QY 90 ----- 90
 DB 2312 AAGGCGAAACTGATCCGATCGAAACGCTCTTTCTCAAAATTAATTGTTATATCGAACT 2371
 QY 90 ----- 90
 DB 2372 CGGCGAAAGCCAAACACAGAGAGCTTCGAAAATTTGATGTTAAAGCATATATACACT 2431
 QY 90 ----- 90
 DB 2432 TTACGAAATGAGCTCTGCAAAAGAAATACATCAACATGTTTACTCTGTTTATCGAA 2491
 QY 90 ----- 90
 DB 2492 GAAAGTTTAATCGATTTGCTTTATCTGCAAACTTGATCCGCGTTTACAGCTTGANT 2551
 QY 90 ----- 90
 DB 2552 TCAGTTCGCTTCAGATGTTAAATCTCAGACGTTGATGATGATGATTTGCTACTCCA 2611
 QY 90 ----- 90
 DB 2612 CTCGAATTCGAAATGTGATAAATTTGGAAGCGAGCAACTCTATCATCGGCCAAACAG 2671
 QY 90 ----- 90
 DB 2672 TAAGAATTTTGAGATTAAAGTTCAAGTTCTCTGCAAAAATCAAAACCGGTGAGAAAT 2731
 QY 90 ----- 90
 DB 2732 TTGCTATGCGGTAGTAGATATCAATATACATCTGCGAAACGAAATTTCTGGCAA 2791
 QY 90 ----- 90
 DB 2792 CCGTATCGTCTTAATCTAGTATCGCTGCGACATATCATATCAGTGAAGGAGTTT 2851
 QY 90 ----- 90
 DB 2852 GGGCTTTGTTGGCTCGATGCTTCAGAGAACAAATTAAGTGTGTTAAGCGGCTAAC 2911
 QY 90 ----- 90
 DB 2912 TCTACCAATCAGAAACGCTTATTCGAGAACCATGTTGTTCTCAATTCCTCCATCCCTA 2971
 QY 90 ----- 90
 DB 2972 CGTACAATCGGGCTTTCCCATTTAGTTCCTTAGGAAGTTGACTTTCACAAACCATTC 3031
 QY 90 ----- 90
 DB 3032 TGGGATTTGTAAGTTGACGTGACGAGCAAGCAAAATATTCATTTGACGAGACTCAAAATTT 3091
 QY 90 ----- 90
 DB 3092 CAATGCTTCCGGGCTTTAGTACTTAATGAGAGTTCAGGAGTTGTTATTTGGGAC 3151
 QY 90 ----- 90
 DB 3152 TCAGCAATGATTAAGAAGAGAAATGTTTTCAGAAAGTATTAATGTTCTTCGCGCTTG 3211
 QY 90 ----- 90
 DB 3212 GGGACTACATCTTCTCTACAGACAAATCTGATGTTTAGGCAAGAGAACTATTACAA 3271
 QY 90 ----- 90

Db 3272 CTGATTAAATAGAGAAAGAGATGGTTCTCCTGGGCAATTTATTGTTAGAA 3331
 Qy 90 ----- 90
 Db 3332 AGCAATTGATATGAATTGGTCTGAGTGTGAAATTCTAGTAGTTGTGTCTT 3391
 Qy 90 ----- 90
 Db 3392 AGTTTCCTGATGTTTATGCTTATCTTGGCACTATCTGGTAGTATCGCTTCTT 3451
 Qy 91 ----- AlalyProCyalySerLeuTriValGlyIleGly 103
 Db 3452 ATGACACCTTTCTGTTGTCAGGCAAACTTGTAAGAGCTGATGGGCTGGAATCGC 3511
 Qy 104 ProAsnValSerIysAspAspLeuGluGluGlnPheSerIysPheGlyIleGlyAsp 123
 Db 3512 CCTAATGCTTCAGAGATGACCTGAGAGAGAGTTGAGCAAGTTGGGAAATCGAGAT 3571
 Qy 124 PheArgPheLeuArgIleArgIleThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
 Db 3572 TTATGGTTTCTCAGAGAAAGCAAGACAGCTTTCAATTATATGATGATGATGCT 3631
 Qy 144 LeuGlnAlaIysSerMetLeuGlyIlePheMetGlyIleSerPheLeuArgValAspPhe 163
 Db 3632 TTACAGGCTAAGAGCAATGAATGAAAGCCCTATGGGTGTAGCTTTTGCCTGTGATTTT 3691
 Qy 164 LeuArgSerGlnAlaPheProIysIys 171
 Db 3692 CTCGGTCCACAGGCGCAAAAAA 3715
 RESULT 6
 ID ABA95169 standard; DNA; 349 BP.
 AC ABA95169;
 XX
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) coding fragment.
 XX
 KM FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigidia;
 KM flowering locus C; transgenic; gene; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200212518-A2.
 PD 14-FEB-2002.
 PF 02-AUG-2001; 2001WO-US024427.
 PR 03-AUG-2000; 2000US-0222550P.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
 DR WPI; 2002-227160/28.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 PT induction gene (FPA) gene which controls flowering time in plants, useful
 PT for altering flowering time of plants and for downregulating flowering
 PT locus C (FLC) mRNA activity.
 XX
 PS Claim 5; Page 37; 39pp; English.
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, *frigida* (FRI) and *FLC* which confer late-flowering

CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FPA gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene coding fragment
 XX
 SQ Sequence 349 BP; 92 A; 68 C; 81 G; 108 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,12e-16 Length: 349
 Score: 466.00 Matches: 92
 Percent Similarity: 94.95% Conservative: 2
 Best Local Similarity: 92.93% Mismatches: 3
 Query Match: 9.71% Indels: 2
 DB: 6 Gaps: 1
 US-09-920-705-3 (1-901) x ABA95169 (1-349)
 Qy 1 MetAlaLeuSerMetIysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
 Db 1 ATGGCGTTATCTATGAAACCATTCAGAGCCGATGATTCGGTTTCCAGTCAAAACATCTT 60
 Qy 21 TTPValGlySerLeuThrProGluThrThrGlnSerAspLeuThrGluLeuPheGlyArg 40
 Db 61 TGGGTGGTAGCTTACCCCGAGACGACAGATGATGACGAGGATGTTGTTGAGAGA 120
 Qy 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 Db 121 TACGCGATATGATGATGAAATCAGCGTATCTTCAGAGGCTTTCGCTTATATATCTAC 180
 Qy 61 ArgHisValGluGluAlaValAlaAlaIysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 181 AGACATGTGGAGAGAGAGCGTGCAGCCAAAGAGGCTCTTCAAGAGCAAAATTTGAATGA 240
 Qy 81 SerGlnIleIysIleGluTyrAlaArgProAlaIysProCysIysSerLeuTriVal 99
 Db 241 AGTCAAATTAAAGATGAAATACGACGACCGGTT-----TGTCTTATCTATATCTT 291
 RESULT 7
 ID ABZ21692 standard; CDNA; 6836 BP.
 AC ABZ21692;
 XX
 DT 27-FEB-2003 (first entry)
 DE Human RBM15-MKL1 fusion protein encoding CDNA SEQ ID NO:1.
 XX
 KM Human; RBM15; RNA binding motif protein 15; megakaryoblastic leukaemia 1;
 KM MKL1; fusion protein; acute megakaryoblastic leukaemia; AMKL; cytostatic;
 KM c(1; 22) chromosomal rearrangement; gene therapy; chromosome 1p13;
 XX
 OS Homo sapiens.
 XX
 PI Key Location/Qualifiers
 FT CDS 84..5735
 FT /*tag= a
 FT /product= "RBM15-MKL1 fusion protein"
 XX
 PN WO200288309-A2.
 PD 07-NOV-2002.
 XX
 PR 23-APR-2002; 2002WO-US012797.
 XX
 PR 27-APR-2001; 2001US-0286910P.
 XX
 PA (SUTD-) ST JUDE CHILDREN'S RES HOSPITAL.
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
 XX

PI Morris SW, Ma Z, Hitzler JK;
 XX MPI: 2003-103455/09.
 DR P-PSDB; ABP56065.
 PT New RNA-binding motif protein-15 (RBM15)-megakaryoblastic leukemia-1
 PT (MKL1), MKL1-RBM15-S and MKL1-RBM15-S+AE fusion protein, useful for
 PT identifying agents useful for treating patients with acute
 PT megakaryoblastic leukemia.
 XX
 PS Claim 2; Page 53-62; 109pp; English.
 XX
 CC The present invention describes an RNA-binding motif protein-15 (RBM15)-
 CC megakaryoblastic leukemia-1 (MKL1) fusion protein, a MKL1-RBM15-S fusion
 CC protein, and a MKL1-RBM15-S+AE fusion protein associated with acute
 CC megakaryoblastic leukemia (AMKL). Also described: (1) an antibody that
 CC specifically binds to the RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
 CC fusion proteins; (2) a non-human transgenic animal that has been altered
 CC to express a gene encoding a RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
 CC fusion protein; (3) identifying an agent capable of binding to a RBM15-
 CC MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion protein; (4) detecting the
 CC t(1;22) chromosomal rearrangement associated with AMKL; and (5) screening
 CC for agents capable of (selectively) inhibiting the activity of a fusion
 CC protein arising from the t(1;22) chromosomal rearrangement associated
 CC with AMKL. The fusion proteins have cytostatic activity and can be used
 CC in gene therapy. The RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion
 CC proteins and nucleotide molecules are useful for designing and preparing
 CC agents that specifically inhibit the expression of the RBM15-MKL1 or MKL1-
 CC RBM15 genes in cells for therapeutic and other purposes. The transgenic
 CC animals are useful for identifying and testing carcinogenic or
 CC therapeutic compositions. The methods are also useful for detecting the
 CC t(1;22) chromosomal rearrangement associated with AMKL, or for
 CC identifying agents useful for treating patients with AMKL. The antibodies
 CC can be used to selectively kill cells expressing RBM15-MKL1, MKL1-RBM15-
 CC S, or MKL1-RBM15-S+AE fusion protein. RBM15 is located to chromosome
 CC 1p13, and MKL1 is located to chromosome 22q13. The present sequence
 CC encodes the RBM15-MKL1 fusion protein of the present invention
 XX
 XX Sequence 6836 BP; 1473 A; 2089 C; 1963 G; 1311 T; 0 U; 0 Other;
 SO
 Alignment Scores:
 Pred. No.: 2,11e-08 Length: 6836
 Score: 336.50 Matches: 237
 Percent Similarity: 34.37% Conservative: 127
 Best Local Similarity: 22.38% Mismatches: 371
 Query Match: 7.01% Gaps: 324
 DB: 8 Indels: 48
 US-09-920-705-3 (1-901) x AB221692 (1-6836)
 QY 7 ProPheArg-----AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
 DB 1152 CTTTCGAGAGAGTGATGAGATTTCACCGAGGATATCAGACGTAACCGGACGCTC 1211
 QY 21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
 DB 1212 TTCTTGCGCAACCTAGACATCACTGTAACGAGAGTATTTAAGAGGCGCTTGATGCGC 1271
 QY 41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
 DB 1272 TTGGAGCTATCAAGAGATGATATCAAGAGCGCTTCGCGGCGAGACTAGTACTTAC 1331
 QY 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaYsgIuAlaLeuGlnGly 75
 DB 1332 GCGTTCTCAAAATTTGAGAACTTAGATATCTCACCGGCGCAAAATTAAGCAATGTGCGC 1391
 QY 76 AlaAsnLeuAsnGlnSerGlnIleValIleGluTyrAlaArgProAlaLysProCysIys 95
 DB 1392 AAAATTATATTTGGATATCTCTCAATATGTTATGTAA---GCTACACCCACACACC 1448
 QY 96 SerLeuTyrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGlnGluPhe 115
 DB 1449 CGCTCTCGGGTGGAGGCGTGGGACCTTGGGTTCTCTGCGCCCTGGACAGAGATT 1508

QY 116 SerLysPheGlyLysIleGluAspPheArgGluArgLysThrAlaPheIle 135
 DB 1509 GATGATTGGTGCACCATACGACCATATGACTACGAAAGGTAGATTGGGCAATATATC 1568
 QY 136 AspTyrTyrGlyLeuTyrAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysPromet 154
 DB 1569 CAGTATGAAGCCTGTGATGACGAGCATGCTGCTGAGACCATATGGGGCTTCCACTT 1628
 QY 155 GlyGly-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGlu 172
 DB 1629 GGTGGCCAGATGACGAGCTTAGAGTACTTGGCCAGACCGAATCATGCTTACACAGAG 1688
 QY 173 GlnTrp-----AlaGlySerTyrAspAsn 180
 DB 1689 CAGTATCTGACAGCTCTGCTGCTGACTTATATGACCTGTGTACAGATGCTTTTGGACAT 1748
 QY 181 Arg-----AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
 DB 1749 CGGGACACAGACCTTTGAGGGGTGCTGGGATAGACACCA-----CCCTTACTATAC 1802
 QY 196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProThr 215
 DB 1803 AGAGATCGTATAGGAGACCTTATCTGACTGTAT---TGGGTGCCACCCACACCCCA 1859
 QY 216 AlaThrGlnCysAsnAspGlnGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 225
 DB 1860 GTCCGA-----GAGCCGACACTCGGACTGACCTTCTGCTGCTGTAT 1907
 QY 236 GluArgValLysSerTyrProSerArg-----AsnPheAlaLeuValGluPheArgSer 253
 DB 1908 GAGCCATGATATGCTTAATATGACGAGCGGATGTGTGCTTACCGGACAGAGGT 1967
 QY 254 AlaGluAlaAlaArgLysGlyGluGlyLeuGlnGlyArgLeuPheAsnProArg 273
 DB 1968 GATGAGATCTGCGCCAGACAGAGAC-----CAGCTTAG 2003
 QY 274 IleValIleMetLysSerAsnAspGluLeuProProGlnGlnAsp-----Asp 289
 DB 2004 AAGCGA-----AGCTGCTCGAGAGAGTGGAGAGCTGACATCTGAT 2045
 QY 290 ThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspProSer 309
 DB 2046 AGGTCTCTGAGAGTACCGCCACGAAACG-----CAC 2081
 QY 310 CysValSerSerPro---HisSerThrGlyIleProGlySerMetArgProLeuArgGly 328
 DB 2082 TGGCTCTCTCTCTGACCGAGTCCAGAAATGACGATGACCGGATGTTACACAG 2141
 QY 329 ThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsn 348
 DB 2142 GACAAATGATCGATCT-----TCCCGTCTCTCTTGAAGGCCCTCT 2183
 QY 349 TrpArgArgProSerLysAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIle 368
 DB 2184 -----CAATCAAGACGAGCGAGGTAGTTG----- 2210
 QY 369 LeuProSerProAlaGlnGlyThrArgArgPrometArgSerAsnProAspSerTrpGlu 388
 DB 2211 -----GAGAAAGACGAGGTGACAAAGCAGAC---CGTAAAACTCTGATAGCTGAA 2261
 QY 389 GlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerVal 408
 DB 2262 -----CGAGATAGAGAACACCGGACCACTGCTCCACT 2294
 QY 409 AspGlyPheThrPrometGlyValAspGluArgSerPheGlyArgGly-----Ser 425
 DB 2295 GAGGGAAGAAAGCCTCTGAAAAAGAACCGCTCTGATGGAAGTGCACCTAGCACACAC 2354
 QY 426 ValAlaAlaArgProIleArgLysPro----- 434
 DB 2355 ACTGCTCTCTCAAGCTGAAGTCCCGTCCAGAAAACAGATGGGGGACAGGCCCTGTG 2414

```

QY 435 -----ProaSerAspHisIleThrArgLysMet----- 444
DB 2415 GCATCAGCCTCTCCCAAACTCTGTTGGCTGGAGGGGCAATGCTTCTACTGAAGAACAGC 2474
QY 445 -----IleAlaIleGlyGlyThrProValCysCysAlaArgCys 457
DB 2475 AACTTCTCTTCAACATGATCATCTGTTCAGGGTGACCTCCAGTGGGTACTACTCTTCT 2534
QY 458 Val-----PrometGlyIleGlyIleGlyIleThrLysLeuProGluValValAsnCys 474
DB 2535 GTGAGGGGTTCACATGTGAGGCAAAAGTGCCAGCTCAAGTCACTGACCTCTCCGTTTG 2594
QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaIleHisIleTyrAlaValAla-----IleGly 492
DB 2595 GAC---CAGCCCAAGTTGATGAAAGTAACTCGACGATCAAAAGTAGCAGGGCCCAATGGT 2651
QY 493 CyeGluIleValPhePheValProAspArgLysGlu----- 504
DB 2652 TATGCCATCTTTGGCTGTGCTGGAGAGTTTGACAGCCGGTCTCTCTTCCTCAGCT 2711
QY 505 -----AspPheAlaSerTyrThrGlu-----PheLeuArgTyrLeuSer 517
DB 2712 GCATCAGACACTGCTCTTACTCTACAGAGCCACTTAGAACCTTGTCTTATTAAAG 2771
QY 518 SerIleAspArgAlaGlyValAlaIleLysLeuAspGly----- 530
DB 2772 CAAGAAGCAGCAGCCGGGTGATCAGCTCTCTGTGGGGGCAACAAGACAGAAAGAAC 2831
QY 531 ---ThrThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGln----- 547
DB 2832 ACCGGGGTCTTATGCTTCCACCTTGAGTTCTCCAGACAGTCTCTGATTCCTCCT 2891
QY 548 -----ValThrArgGlnIleArgLeuTyrGlyValValLeuLysLeuProPro 564
DB 2892 GCCAAGCACTGGCCAAATCTGAAGAAGATTACTGTGATCATCTTC-----CGT 2945
QY 565 AlaValProValThrAlaSerTyrArgGlnLysSerGlnSerAsnProLeuHisIleTyrMet 584
DB 2946 GCTTTGAAGAAGTCCAGCCGCAATTCATGACAGAGAAAGAGC-----TTG 2990
QY 585 AspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgLysAsn 604
DB 2991 GAGCGGGCCAGG-----ACAGAGGAC 3011
QY 605 Tyr-----IleArgGlyAlaProGluHisIleuThrAlaIleSerLysProSer 620
DB 3012 TATCTCAAGCAAGAAATTCCTCCCGCGAGAGATCGAAGCTGTCCAGATGCACATT 3071
QY 621 ValSerGluProLeuArgIleProAsnAlaAlaIleProGlnAlaGlyValSer----- 638
DB 3072 TTGGAAGACACTCGGCTGAGCATCTCCAGGCCAGACAGCTGAAAGTGAAGAGAGCC 3131
QY 639 ---LeuThrProGluLeuLeuAlaThrLeuAla----- 648
DB 3132 AGACTGACCGATGACCTCAATGAAGAAATTCACAGAGGCTGGCCCATGAGCTGTGG 3191
QY 649 -----SerIleLeuProAlaThrSerGln----- 656
DB 3192 GAGAAAGACATCTCTCTGTAGTCCAGCTGAGAGAGACCATCATTTGGGCCAGAGTG 3251
QY 657 -----ProAlaIleProGluSerHisGlnProMetSerGlyProSerThrValValSer 674
DB 3252 AACTATCCCAAGTAGCAGACAGCTCTCTCGATGAGAGCAGCAGCATGCTTATCC 3311
QY 675 ThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyValAlaProSerGlnAlaIleTyrLysArg 694
DB 3312 CCCGAGCAAGCTGCGAGCATATGATCCAGGGGTCTGTGCTGTCACCCCTGAGAGCCGGA 3371
QY 695 GlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyr 714
DB 3372 GTACAGGAACCACTGCTAGTGCACACTGTGATCCCAACCAAGGTGTGTCTAA--- 3428
QY 715 ThrProAlaGlyGln-----LeuPro 721

```

```

DB 3429 CTTCCGATGGCGCGGATTCAGAGAAATGCTTTTCTGGCAGACAGCCTCTTGCT 3488
QY 722 ProProSerAspArgTyrProProAlaSerAsn-AsnProAsnTyrThrSerGlyMetVal 741
DB 3489 CCCCACCTCTGTCTGCTCTCCAGCTCACCAATGAAACCA----- 3528
QY 741 HisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuPr 761
DB 3529 -----CTATCCCCACTGCC 3542
QY 761 AsnMetProHisAsnAsnTyrSerMetTyrThrGlnLysSerAsnHisProValSer 781
DB 3543 AGTCCACCC-----CCACACTCATTTAAG 3566
QY 781 r-GlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyP 801
DB 3567 CAAGCCA---ACCAAGTCTGCGAGTGAAGAGTACAGCGCAGCAGAGAGCCAGAGG 3623
QY 801 ro-IleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
DB 3624 CTGAAGCCAAAGGTGAGAGAGCTCAAGTACAC-----CAGTAC 3662
QY 821 LeuAsnProSerGlnPheGlnAlaIleMetGlnProProAlaAspLysAlaAsnLeuGln 840
DB 3663 ATCCCCCGGACCAAGACAGAGACAGAGGAGGAGCAGCCCATGAGACTCATCTACGCCAAG 3722
QY 841 ---ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThr 859
DB 3723 ATCTGACAGACAGACAGACTCTTCTCTCAG----- 3752
QY 860 ThrAspGlyGlyValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAsn 879
DB 3752 ----- 3752
QY 880 LeuLeuLeuGlnIleGlnGlnLysGlnGlnGln----- 891
DB 3753 ---CTGCAATCTCTAACACAGCAGCAGCAGCAGCAGCAGCACTACAGGCCATCTCG 3806
QY 892 -----SerSerGlyTyrPro 896
DB 3807 CTTGCCCGCCAAAGTACAGCAGCGAGGCTCTGGAGAGCAGCGGAGACCC 3857
RESULT 8
AAZ51263
ID AAZ51263 standard; cDNA; 2190 BP.
XX
AC AAZ51263;
XX
DT 06-JUN-2000 (first entry)
XX
DE Human RNA-associated protein-14 (RNAAP-14) encoding cDNA.
XX
KW RNA-associated protein; RNAAP; human; clone 1748626; cytosolic;
XX immunosuppressive; antiinflammatory; keratolytic; neuroprotective;
XX antifibrotic; hepatoprotective; antiproliferative; virucide; anti-HIV;
XX antiallergic; antineumatic; antiarthritic; ophthalmological; autoimmune;
XX antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
XX acutinic keratosis; buritis; arteriosclerosis; arthritis; cancer;
XX hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
XX mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
XX allergy; rheumatoid arthritis; parasitic infection; ss.
XX
OS Homo sapiens.
XX
FH Key location/qualifiers
FH CDS 164..1855
FH FT /*tag= a
FH FT /product= "Human RNA-associated protein-14"
FH FT /note= "Derived from STOMTUT02 library"
XX
PN W020001171-A2.
XX

```


PD 02-MAR-2000.
 XX 20-AUG-1999; 99MO-US019361.
 XX 21-AUG-1998; 98US-0097550P.
 PR 12-JAN-1999; 99US-0115639P.
 XX (INCY-) INCYTE PHARM INC.
 XX Hillman JL, Yue H, Tang YT, Corley NC, Guejler KJ, Gorgone GA,
 PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y,
 PI Shih LT, Yang J, Lu DM;
 DR WPI; 2000-237651/20.
 DR P-PSDB; AAY70233.
 XX Human RNA-associated proteins useful in diagnosing, treating and
 PT preventing cell proliferative, autoimmune, inflammatory and infectious
 PT disorders.
 XX
 PS Claim 9, Page 116-117, 123pp, English.
 CC The present sequence is the cDNA encoding human RNA-associated protein-14
 CC (RNAAP-14), identified in incyte clone 1748626, derived from STOMT02
 CC library. It is expressed in reproductive, nervous, gastrointestinal, and
 CC haematopoietic/immune tissues. It has cytosolic, immunosuppressive,
 CC antiinflammatory, antiarteriosclerotic, hepatocytic, keratolytic,
 CC neuroprotective, antiparasitic, anti-HIV, antiallergic, antirheumatic,
 CC antiviral, antitubercular, ophthalmological and antimicrobial activity.
 CC RNAAP antibodies are useful for diagnosis of diseases associated with
 CC altered expression or activity of RNAAP. It is used to treat cell
 CC proliferative, autoimmune, inflammatory and infectious disorders, like
 CC actinic keratosis, burns, arthritis, arteriosclerosis, artherosclerosis,
 CC cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease
 CC (MCTD), psoriasis, primary thrombocytopenia and cancer, HIV, allergies,
 CC rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and
 CC parasitic infections
 XX
 SQ Sequence 2190 BP; 515 A; 616 C; 633 G; 426 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4,22e-07 Length: 2190
 Score: 302.00 Matches: 157
 Percent Similarity: 33.67% Conservative: 78
 Best Local Similarity: 22.49% Mismatches: 232
 Query Match: 6.29% Gaps: 231
 DB: 3 Gaps: 27
 US-09-920-705-3 (1-901) x AAZ51263 (1-2190)
 QY 11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
 DB 170 GAGGATGACCAAGCGGCGGCGCAACCTTCATTCATGTAACCTGAGCACACAGCATCT 229
 QY 31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
 DB 230 GAGGTGAGCTGCGAAGGCGCTTCGAGAAATATGCGCATCGAAGAGGTGTCATCAAG 289
 QY 50 TyrSerSerArgGly-----PheAlaPheIleTyrTyrArgHisValGluGlu 65
 DB 290 AGGCGTGGCCGTGCGGCGGCGGCGGCTATGCTTCCTCAAGTTCGAAACCTGAGCATG 349
 QY 66 AlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleValIle 85
 DB 350 GCCCATAGGAGCTAAAGTGGCGCATGTCCGGCGGAGTGTGTCGCAACCCCATTAAGATA 409
 QY 86 GluTyrAlaArgProGluAlaAspProCysValSerSerLeuTrpValGlyGlyIleGlyProAsn 105
 DB 410 GGCATATGGCAAG--GCCAACCCCAACCACTGCTTCTGGGTGGGTGGCTGGAGACCTAAC 466
 QY 106 ValSerTyrAspAspLeuGluGluGluGluPheSerTyrPheGlyLysIleGluAspPheArg 125
 DB 467 ACCTACCTGGCGGCTGCGGCGGCGGAGGTTTGAACCGCTTGGGAGACANTTGGACCATTTGAT 526

QY 126 PheLeuArgGluArgIleThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
 DB 527 CAGCTAAAGAGATAGCTTTGCTCTATATCATGATACGAGACCTTGAGCCAGGCC 586
 QY 146 Ala---LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
 DB 587 GCTTGCTGCTAAATATGAGGGGTTTCCCTTGCTGAGACGACCGAGGCTCCCGTGGAT 646
 QY 163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAsnArgAsn 182
 DB 647 TTGCGCAAGACAGAG----- 661
 QY 183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
 DB 662 -----GAGACTCGGTACCCCGACAGCATAC----- 685
 QY 203 GlnProSerTyrValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
 DB 686 CAGCCTTCGCCACTC-----CCTGTCATTAAGAGTCTGCTCACAGATGGA 730
 QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
 DB 731 TACACCGGCGACCGCAAC-----CTGACGCGCGACTGTGGGGGAGACAGACGCCCA 784
 QY 243 -----SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
 DB 785 CACCTTCTGTACTACAGACCGAGACCGGACTTTTGGAAAGGAGCTGACACGCCCACT 844
 QY 256 GluAlaArgGlnCysLysGluGluGlnGlyArgLeuPheAsnAsnProArgIleLys 275
 DB 845 AAAAGCTCGACCGCGGAAACAGCTTGAGGC----- 877
 QY 276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
 DB 878 -----TACAGTCCG 886
 QY 296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProH 315
 DB 887 TCAGTGGCGAGCGGAGGTGAGCGTGGGGGCGAGATGAGACCGGTGTTGGCCCAAG 946
 QY 315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
 DB 947 C-----CCTGGG-----AAGNAGCGCGAA 967
 QY 335 nglyAlaGluTyrAsnAspValValGlyLysGluPro-----AsnTrpArg 350
 DB 968 CGGAGAAAGCTTCCAGTACCGGTGGAGAGACACCATTCACATATGAGAAACGAGT 1027
 QY 350 gatGProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----- 366
 DB 1028 AGGACCAAGGAGAGTGGGAGAGATCAGACGGGGCTTCGACCGCACCCCTGAGCGCAGC 1087
 QY 367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg----- 378
 DB 1088 CGCAAGAAACACATCTCCAGTGAAGGACCAAGAGTCCAGACGAACTCCCTCAGCAAC 1147
 QY 378 ----- 378
 DB 1148 AGCAGATAGGGGCTGAGGAAGGGGCGACACACACACACAGAGGCTGAGACTCT 1207
 QY 379 -----PrometArgSerAsnPr 384
 DB 1208 TCCCAAGGAGAAAGCAAGACAGACGCGCAATACCGGACCAAGAGCGCGAGCC 1267
 QY 384 GAspSerTrpGluGluTyrAspProAla-GlnLeuValArgLysSerLysArgThrArgA 404
 DB 1268 AAGCCTCTGAAAGAGCAAAACACAGACCAAAAGCTTAAGAAATTTTCAAGATACGCT 1327
 QY 404 rGAspGlySerValAspGlyPheThrPrometGlyValAspAlaLys----- 419
 DB 1328 CAGACACTACAGCTGGGTTGAGATGAGGCTTCTGGTGTGTAATAAACAGCTGTTCCACAG 1387

```
QY 420 -----SerPheGlyArgGlySerValAlaIala----- 428
DB 1388 TCTATGCTATCTCTAGAGGGGAGACAGGGGGTGTATCGACTCTCTCAAGACCACT 1447
QY 429 -----ArgProIleArgGlyPro-ProAsp 436
DB 1448 TCTGGAGCAAGCTGACCCAGCTGAAGATCGCCAGCGCTTGACACTGACCAAGCCCAAG 1507
QY 437 SerAspHisIleIleTyrPheGlyMetIleAlaGlyGlyThrPro----- 451
DB 1508 CTTGACGAGTTCACACGACGCTC-----AAGCAGGGGAGCCCAAGCGCTATGCGGTC 1561
QY 452 ValCysSerAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
DB 1562 CTCTTAGCCACCGACGACACCCCAAGTGGCTTGACACTAG--GGATGCCACAGTA 1618
QY 472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
DB 1619 GAGCCCGGTCTGACAGGCGGCTTCTCAGGACCTG----- 1654
QY 492 GlyCysGluIleValPhePheValProAspArgGluAspPheAlaSerTyrThrGlu 511
DB 1654 ----- 1654
QY 512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp----- 528
DB 1655 ---GTCTCTTACTTGAACAGAAAGAGCGCGGAGGGGTATCGCTTGCCAGTGGGGGG 1711
QY 529 -----AspGlyThr---ThrLeuPheLeuValProProSerAspPheLeuThr 543
DB 1712 TCCAAAGGCGAAGACCGACAGCATGCTCTACGCTCCACCTCCAGCACTTTCCAG 1771
QY 544 AspValLeuGluIleValThr-----ArgGluGluArgLeuTyrGly 556
DB 1772 CAGTACTCTCAGTCAGACACTTAAGACATTGGCGCACTTGAGAGAACACATGGTATG 1831
QY 557 ValValLeuLys-----LeuProProProAlaValProValThrAlaSerTyrArg 573
DB 1832 GTCATCGTCAGACACACTGCTGACCAAGCTGCTTTCCAGCGCTCANTGTTGTGTA 1891
QY 574 GlnGluSer-----GlnSerAsnProLeuHisTyr 583
DB 1892 CAAMAAGCAGTATTATTAATCTGATCCCTCTCTACCTTACCACTTT 1939

RESULT 9
ACN37956
ID ACN37956 standard; cDNA; 3127 BP.
AC
ACN37956;
DB 18-NOV-2004 (first entry)
XX Tumour-associated antigenic target (TAT) cDNA DAA324333, SEQ ID NO:1186.
DE
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
OS Homo sapiens.
XX
XX PN WO2004030615-A2.
XX
XX PD 15-APR-2004.
XX
XX PF 29-SEP-2003; 2003WO-US028547.
XX
XX PR 02-OCT-2002; 2002US-0414971P.
XX
XX PA (GETH ) GENENTECH INC.
```

```
XX Wu TD, Zhang Z, Zhou Y;
PI
PI MPI; 2004-347921/32.
DR
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. Breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 1; SEQ ID NO 1186; 7273bp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid, an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention
XX
XX Sequence 3127 BP; 629 A; 1001 C; 966 G; 531 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 6.2e-07 Length: 3127
Score: 302.00 Matches: 157
Percent Similarity: 33.67% Conserved: 78
Best Local Similarity: 22.49% Mismatches: 232
Query Match: 6.29% Indels: 231
DB: 13 Gaps: 27

US-09-920-705-3 (1-901) x ACN37956 (1-3127)
QY 11 AspaSerGlyPheGlnSerAsnLeuTyrValGlySerLeuThrProGluThrThr 30
DB 1112 GAGATGACCAAGCGGCGCACCGCACTCTTCAATGATGACCTGACCAAGCGTATCT 1171
QY 31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
DB 1172 GAGGTGAGACTCCAAAGGCGCTTCGAAATATGCAATCTGAGAGGTGATCATG 1231
QY 50 TyrSerSerArgGly-----PheAlaPheIleTyrTyrArgHisValGluGlu 65
DB 1232 AGGCTGCGCGGTGCGAGGCGGTGCTTCTCTCAAGTTCAGAACTGGAATG 1291
QY 66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
DB 1292 GCCATAGGGCTAAGTGGCCATGTGCGGCGAGTGTGTCGCAACCCCATTAAGATA 1351
QY 86 GluTyrAlaArgProAlaLysProCysLysSerLeuTyrValGlyGlyIleGlyProAsn 105
DB 1352 GGTATGGCAAG---GCCAACCCACCACTGCTTCTGGGTGGGTGGCTGGAGACTTAC 1408
QY 106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
DB 1409 ACGTCACTGGCGGCTTGCGCCGAGAGTTGACCGCTTTGGAGACATTCGACCATGAT 1468
QY 126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAlaLeuGln 145
DB 1469 CACGTCAAGAGATGACTTGTCTTATATTCAGTACGAGAGCTTGGACGCGCAGGCC 1528
```

```

QY 146 Ala--LysSerMetAenGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
Db 1529 GCCTGCTGCTAAATGAGGGGTTTTCCTTGGGTGACACGACGAGCTCCCGCTGGAT 1588
QY 163 PheLeuArgSerGlnAlaProLysGluGlnTrrAlaGlySerTyrAenAArgAen 182
Db 1589 TTGGCCAAAGCAGAG----- 1603
QY 183 GlyAenMetAenHisLysPProGlnTyrProHisSerTyrGluAenPheLysGlyAenPVal 202
Db 1604 -----GAGACTCGGTACCCCGACAGTAC----- 1627
QY 203 GlnProSerLysValLeuTrrIleGlyPheProProThrAlaThrGlnCysAenAenGlu 222
Db 1628 CAGCCCTCGCCACTC-----CCTGTGATTATGAGCTGCTCAGATGCA 1672
QY 223 GlnIleLeuHisAenAlaMetIleLeuPheGlyGlnIleGluArgValLysSerTyrPro 242
Db 1673 TACACCCGCGACCGCAAC-----CTGACGCGGACTGGTGCGGACAGAGCCGCCCA 1726
QY 243 -----SerArgAenPheAlaLeuValGluPheArgSerAlaGlu 255
Db 1727 CACCTTCTGACTACAGACCGACCGGACTTTTGGAGGGGAGCTGGACCCAGT 1786
QY 256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAenAenProArgIleLys 275
Db 1787 AAAAGCTCTGACCGCGCAACAGCCTTGAGGCG----- 1819
QY 276 IleMetLysSerAenAenAenGluLeuProProGluGlnAenAenPThrSerPheTyrSerGly 295
Db 1820 -----TACAGTCCG 1828
QY 296 MetLysArgSerArgThrAenP-MetPheAenAenAenPProSerCysValSerSerProHis 315
Db 1829 TCAGTGGCGACCGCGAGGTGTCGCTTGGGCGGCGAGATGAGACCGCTGTTGCCAAG 1888
QY 315 SserThrGlyIleProGlySerMetAenPProLeuArgGlyThrAenGluArgSerTyrAs 335
Db 1889 C-----CCTGGG-----AAGAGAGCGCGAAA 1909
QY 335 nglyAlaGlyTyrAenAenAenPValGlyLysGluPro-----AenTrrArg 350
Db 1910 CGAGAGAAGCCTTTCAGTAGACCGTGGAGAGACCAACCATTCACATATGAGAACGAGT 1969
QY 350 garPProSerAlaAenGlyThrGlyIleLeuProSerProThrGlyPro----- 366
Db 1970 AGGACCAAGGCGAGTGGCGAGCAGTCAAGCGGGGCTCCGACCGCACCTTGAGCGCAGC 2029
QY 367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg----- 378
Db 2030 CGCAAGAGAACCACTCCAGTAGAGGACCAAGAGATCCAGACCACTCCTCAGCAAC 2089
QY 378 ----- 378
Db 2090 AGCAGACATGGGGCTGAGAAACGGGGCCACACACACACAGAGGTGAGACTCT 2149
QY 379 -----ProMetArgSerAenP 384
Db 2150 TCCCAACGGGAAGCAAGCAAGACAGCGACGCAATCACCGGACCAAGAGCGCGAGCC 2209
QY 384 aaPSPSerTrrGluGlyTyrAenPProAla-GlnLeuValArgGluSerLysArgThrArg 404
Db 2210 AAGCCTCTGAAAGGCCAAACACGAGACCAAAAGCTGAAGATCTTTCAGAGTACGT 2269
QY 404 rgAenPArgSerValAenGlyPheThrProMetGlyValAenGluArg----- 419
Db 2270 CAAGACTACAGCTGGGTGGATGGGCTTGTGTGTAAGAAACAGCTCTCCACAG 2329
QY 420 -----SerPheGlyArgGlySerValAlaAla----- 428
Db 2330 TCTATGATATCTCAGAGGGGAGCAAGGGGTGATCAGAGTCTCTCAAGACCACT 2389
QY 429 -----ArgProIleArgGlyPro-ProAenP 436

```

```

Db 2390 TCTGGAGAGACTGACCCAGCTGAAGATCGCCAGCGCTTCGACTGACACGCCAAG 2449
QY 437 SerAenHisIleTrrArgGlyMetIleAlaLysGlyThrPro----- 451
Db 2450 CTTGACGAGATGACACAGCAGCATC-----AAGCAGGGAGGCCCAACGGCTATGCGCTC 2503
QY 452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
Db 2504 CTTTACGCCACCGAGCAACCCCGAGTGGCTTGGACTGAG--GGATGCCACAGTA 2560
QY 472 ValAenCysSerAlaArgThrAenPLeuAenMetLeuAlaLysIleTrrAlaValAlaIle 491
Db 2561 GAGCCCGGTCTGACAGAGGGGCTTCCAGGAACCTC----- 2596
QY 492 GlyCysGluIleValPhePheValProAenArgGluGluAenPheAlaSerTyrThrGlu 511
Db 2596 ----- 2596
QY 512 PheLeuArgTyrLeuSerSerLysAenPArgAlaGlyValAlaLysLeuAenP----- 528
Db 2597 --GTCTCTACTTGAACAGAGAGCGCGAGGGGTGATCAGCTTGGCAGTGGGGGG 2653
QY 529 -----AspGlyThr--ThrLeuPheLeuValProProSerAenPheLeuThr 543
Db 2654 TCCAGAGGACAGAGCGGACAGGACATGCTTACGCTTCCACCTCGCAGCTTTTCCAG 2713
QY 544 AspValLeuGlnValThr-----ArgGlnGluArgLeuTyrGly 556
Db 2714 CAGTACCTCAGTACAGCACTAAGACATGGGCAAGCTGAGACATGAGATGATGATA 2773
QY 557 ValValLeuLys-----LeuProProProAlaValProValThrLysSerTyrArg 573
Db 2774 GTCATGTCAGAGACACTGCTTACGCCAAGCCTGTCTTCCAGCGCTCAATGTTGTGCA 2833
QY 574 GlnGluSer-----GlnSerAenProLeuHisTyr 583
Db 2834 CAAGACAGTATTATTAAATCTGATCCCTCTCTACCTTACACTTT 2881

RESULT 10
AAD53223_4
Continuation (5 of 5) of AAD53223 from base 40001 (Human chromosome 3 p-arm breakpoint
WP Sequence split into 5 fragments LOCUS AAD53223 Accession AAD53223
WP Fragment Name Begin End
WP AAD53223_0 1 110000
WP AAD53223_1 100001 210000
WP AAD53223_2 200001 310000
WP AAD53223_3 300001 410000
WP AAD53223_4 400001 487980

Alignment Scores:
Pred. No.: 2,266-05 Length: 87980
Score: 302.00 Matches: 157
Percent Similarity: 33.67% Conservative: 78
Best Local Similarity: 22.49% Mismatches: 232
Query Match: 6,29% Indels: 221
DB: 8 Gaps: 27

US-09-920-705-3 (1-901) x AAD53223_4 (1-87980)
QY 11 AsnAenSerGlyPheGlnSerAenAenLeuTrrPValGlySerLeuThrProGluThrThr 30
Db 73903 GAGGATGACACCGGGCCAGCGCAACCTTCATTTGTAACCTGAGACCAAGCGTATCT 73962
QY 31 GluSerAenPLeuThrGluLeuPheGlyArgTyrGlyAspIleAenPArgIleThrVal-- 49
Db 73963 GAGCTGAGACTCGGAAGGCGCTTCCAGAAATATGCAATATGAGAGAGGTGATCATCAG 74022
QY 50 TyrSerSerArgGly-----PheAlaPheIleTyrTrrArgHisValGluGlu 65
Db 74023 AAGCTCCCGGTGGCCAGAGGCGGTGCTATGCTTCTTCAGATTCAGAACTGAGACATG 74082
QY 66 AlaValAlaLysGluAlaLeuGlnGlyAlaAenLeuAenGlySerGlnIleLysIle 85

```


XX 24-OCT-2003; 2003MO-US033657.
PF 25-OCT-2002; 2002US-0421080P.
XX 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433318P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 22-MAY-2003; 2003US-0472420P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476632P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hecht K;
XX WPI, 2004-376048/35.
DR P-PSDB; ADS95116.
XX
PT New nucleic acid molecules, useful for treating proliferative disorders,
XX e.g. cancer, and inflammatory, immune, bacterial or viral disorders.
XX
PS Disclosure; SEQ ID NO 74; 220bp; English.
XX
XX The present invention relates to novel human proteins (ADS95043-ADS95123)
CC and DNA sequences (ADS94962-ADS95042 and ADS95124-ADS95204). Also
CC disclosed are murine DNA sequences (ADS95205-ADS95285). The sequences are
CC useful for treating proliferative disorders, e.g. cancer, and
CC inflammatory, immune, bacterial or viral disorders. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published/pct/sequences. Also, Claim 1 of this
CC specification refers to SEQ IDs 419-627 and Claim 14 refers to SEQ ID 210
CC -418, however, the sequence listing comprises 324 sequences.
XX
XX Sequence 2673 BP; 540 A; 883 C; 858 G; 392 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 9.55e-07 Length: 2673
Score: 297.00 Matches: 148
Percent Similarity: 33.90% Conservative: 71
Best Local Similarity: 22.91% Mismatches: 212
Query Match: 6.19% Indels: 215
DB: 13 Gaps: 24
US-09-920-705-3 (1-901) x ADS95035 (1-2673)
Qy 11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
Db 988 GAGATATCCAGCGGCGCGACCGCACCTCTTCACTTGTAACCTGACCAACGAGGTATCT 1047

Qy 31 GluSerAspLeuThrIleuPheGlyArgTrpGlyAspIleAspArgIleThrVal---- 49
Db 1048 GAGGTGAGCTCGAAGGCGCTTCCAGAAATATGGATATCATCGAAGGCTGATCAAG 1107
Qy 50 TyrSerSerArgGly-----PheAlaPheIleTyrTrpArgHisValGluGlu 65
Db 1108 AGGCTTCGCCCTGCGCCAGGCGGCGGTGCTTGCCTTCTCCTCAAGTCCAGAACCGACATG 1167
Qy 66 AlaValAlaValGluValGluGluGluGluGluGluGluGluGluGluGluGlu 85
Db 1168 GCCATAGGCTTAAAGGTGCGCATGTGCGGCCAGATGATGTCGCAACCCCATTAAGATA 1227
Qy 86 GluTrpAlaArgProAlaArgProCysArgSerLeuTrpValGlyIleGlyProAn 105
Db 1228 GGCTATGGCAAG---GCCAACCCCAACCACTGCTCTGTGGTGGTGGTGGTGGTGGTGGTGG 1284
Qy 106 ValSerTyrAspAspLeuGluGluGluGluGluGluGluGluGluGluGluGlu 125
Db 1285 AGCTCACTGGCGGCTTGGCCGAGAGCTTGAACCGCTTGGAGACATTCGAGCATTTGAT 1344
Qy 126 PheLeuArgGluArgTrpAlaPheIleAspTrpTyrGluMetAspAspAlaGlu 145
Db 1345 CACGTCAAGAGATAGCTTGGCTATATTCAATGACAGAGCTTGGAGCCAGCCAGCC 1404
Qy 146 Ala---LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
Db 1405 GCCTGTGCTAAATGAGGGGTGTTTCCCTGGGGGAGCCAGACCGCAGGCTCCGCTGGAT 1464
Qy 163 PheLeuArgSerGlnAlaProLysGluGluGluGluGluGluGluGluGluGlu 182
Db 1465 TTGGCCAAAGCAGAG----- 1479
Qy 183 GlyAsnMetAsnHisLysArgProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
Db 1480 -----GAGACTCGGTACCCCGCCAGCACTAC----- 1503
Qy 203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
Db 1504 CAGCCTTCGCCACTC-----CCTGTGCATTATGACGCTGCTCACACAGATGGA 1548
Qy 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
Db 1549 TACACCCGACCGCAGAAC-----CTGACCGCCGACTGTGTCGCGACAGACGCCCA 1602
Qy 243 -----SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
Db 1603 CACCTTCTGTACTCAGACCGAGACCGGACTTTTGGAAAGGGAGCTGACACCGCCAGT 1662
Qy 256 GluAlaArgGlnCysLysGluGluGluGluGluGluGluGluGluGluGluGlu 275
Db 1663 AAAAGCTTGACCGCCGAAACAGCCCTTGAGGGC----- 1695
Qy 276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
Db 1696 -----TACATGCG 1704
Qy 296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHis 315
Db 1705 TCAGTCCGACGCGAGGTGGTGTAGCGTTGGGGGCGAGATGAGACCGTGGTTGGCCAAAG 1764
Qy 315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
Db 1765 C-----CTGGGG-----AAGAGAGCGCGGAAA 1785
Qy 335 ngIValaGluTrpAsnAspValValGlyLysGluPro-----AsnTrpArg 350
Db 1786 CGAGAGAGCTTTCAGTACCGTGGAGAGACCAACCATTCATATGAGAACCGAGT 1845
Qy 350 gArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----- 366
Db 1846 AGGACCAAGGGCAGTGGCGAGAGTGCAGAGCGGGGCTCCGACGCGACCCCTGAGCGCAGC 1905

QY 367 -GlyLeuProSerProAlaGlnGlyThrArg- 378
Db 1906 CGAAGAGAGAACCACTCAAGAGAGCAAGAGATCCAGCAGCACTCCCTCAGAAC 1965
QY 378 - 378
Db 1966 AGCAGCATGGGGCTGAGAACGGGGCCACCACCACCAAGAGCTGAGACTCT 2025
QY 379 - 379
Db 2026 TCCACGGGAAGAGCAAGACAGCAGCAGCAATCCGAGCCAGAGCGCCGAGCC 2085
QY 384 GAAPSerTyrGlnGlyTyrAspProAla-GlnLeuValArgGlnSerTyrArg 404
Db 2086 AAGCTCTGGAGAGCCAAACACAGAGCAAAAAGATGTAAGATCTTCAGAGTACCT 2145
QY 404 rGAPGlySerValAspGlyPheThrProMetGlyValAspGlnArg- 419
Db 2146 CAGACACTACAGCTGGGTTGAAATGGGCTTCTGGTGTGMAAACAAGCTGCTCCACAG 2205
QY 420 -SerPheGlyArgGlySerValAlaAla- 428
Db 2206 TCTATGATATCTCAAGAGGGGAGCCAGGGGTATATCAGAGCTCTCAAGACCACT 2265
QY 429 - 429
Db 2266 TCTGGAGCAAGCTGACCCAGCTGAGATGCGCCAGCGCTTGCAGTGGACAGCCCAAG 2325
QY 437 SerAspHisLeuTyrPheGlyMetLeuAlaGlyGlyThrPro- 451
Db 2326 CTGAGAGAGTCAACAGCAGCATC- 451
QY 452 ValCysCysAlaArgCysValProMetGlyGlyGlyLeuThrLysLeuProGluVal 471
Db 2380 CTCTTGACCAACCAAGCAGCAACCCCAAGTGGGCTTGGACAGTACG- 471
QY 472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAla 491
Db 2437 GAGCCCGGTCTGAGAGCGGCTTCTCAGGAACCTG- 491
QY 492 GlyCysGlnLeuAlaPhePheValProAspArgGlnAspPheAlaSerTyrThrGlu 511
Db 2472 - 2472
QY 512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp- 528
Db 2473 -GCTCTCACTTGAACAGAGAGCAGCGCGGATATCAGCTTCCACCTGCGAGCTTCCAG 2529
QY 529 -AspGlyThr--ThrLeuPheLeuValProProSerAspPheLeuThr 543
Db 2530 TCCAAAGGAGAGAGCAGCAGCAGCTCTTCCAGCTTCCAGCTTCCAG 2589
QY 544 AspValLeuGln 547
Db 2590 CAGTACTCTCAG 2601
RESULT 12
ADS95197
ID ADS95197 standard; DNA, 2673 BP.
AC ADS95197,
XX
DT 02-DEC-2004 (first entry)
XX
DE Human therapeutic DNA sequence #74, SEQ ID 236.
XX
KW Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Vinorelbine;
KW Gene therapy; Proliferative disorder; cancer; inflammatory disorder;
KW Immune disorder; bacterial disorder; viral disorder; human; gene; ds.
OS Homo sapiens.
XX
PN WO2004039319-A2.

XX 13-MAY-2004.
PD 24-OCT-2003; 2003WO-US033657.
XX
PF 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433318P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 22-MAY-2003; 2003US-0472420P.
PR 09-JUN-2003; 2003US-047609P.
PR 09-JUN-2003; 2003US-0476632P.
PR 08-JUL-2003; 2003US-0485225P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
PI Williams LT, Chu K, Lee E, Heestir K;
XX WPI; 2004-376048/35.
DR P-PsDB; ADS95116.
XX
XX New nucleic acid molecules, useful for treating proliferative disorders,
PT e.g. cancer, and inflammatory, immune, bacterial or viral disorders.
XX
XX Disclosure; SEQ ID NO 236; 220pp; English.
PS
XX The present invention relates to novel human proteins (ADS95043-ADS95123)
CC and DNA sequences (ADS94962-ADS95042 and ADS95124-ADS95204). Also
CC disclosed are murine DNA sequences (ADS95205-ADS95285). The sequences are
CC useful for treating proliferative disorders, e.g. cancer, and
CC inflammatory, immune, bacterial or viral disorders. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp://wipo.int/pub/published_pat_sequences. Also, Claim 1 of this
CC specification refers to SEQ IDs 419-627 and Claim 14 refers to SEQ ID 210
CC -418, however, the sequence listing comprises 324 sequences.
XX
SQ Sequence 2673 BP; 540 A; 883 C; 858 G; 392 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 9.65e-07 Length: 2673
Score: 297.00 Matches: 148
Percent Similarity: 33.90% Conservative: 71
Best Local Similarity: 22.91% Mismatches: 212
Query Match: 6.19% Indels: 215
DB: 13 Gaps: 24
US-09-920-705-3 (1-901) x ADS95197 (1-2673)
QY 11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30


```

Db      988 GAGAGTACACGAGCGGCGACGCGCAACCTTCATTGTAACCTGACACGACGATCT 1047
      31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal 49
      1048 GAGGTGAGGTGGGAGGCGCTTCGAGAAATATGATGATCGAGAGGTGGTCACTCAAG 1107
      50 TyrSerSerArgGly-----PheAlaPheIleTyrTyrArgHisValGluGlu 65
      1108 AGGCTTCGCGGCTGCGGCGGCGGCTGCTTCCTCAAGTTCAGAACTCGACATG 1167
      66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLeuVal 85
      1168 GCCCATAGAGGCTAAGGTGGCCATGTGGCGCGAGTATGTCGCAACCCCATTAAGATA 1227
      86 GluTyrAlaArgProAlaLysProCysLysSerLeuTyrValGlyGlyIleGlyProAsn 105
      1228 GGGTATGGCGAG--CCCAACCCCGACCACTGCTCTGGGTGGTGGCTGGGACCTAAC 1284
      106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
      1285 ACCTCACTGGCGGCTTGGCCGAGAGTTTGACCGCTTGGAGCAATTGGACCATTTGAT 1344
      126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
      1345 CAGCTCAAGAGATAGCTTGGCTTATTCAGTACGAGAGCTTGGACGCGACGCCAGGCC 1404
      146 Ala--LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
      1405 GCTGTGTCTAAATATGAGGGGTTCCTCTGGGTGGACCAAGCCGAGCTCCCGTGGAT 1464
      163 PheLeuArgSerGlnAlaProLysLysGluGlnTyrAlaGlySerTyrAspAsnArgAsn 182
      1465 TTTCGCAAGCGAGAG-- 1479
      183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
      1480 -----GAGACTCGGTACCCCGACAGATAC----- 1503
      203 GlnProSerLysValLeuTyrIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
      1504 CAGCCCTCGCACTC-----CTGTGCATTATGAGCTGCTCAGACATGGA 1548
      223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
      1549 TACACCCGCGACCGCAAC-----CTGGACGCGGCTGTGGGACAGAGAGCCGCCA 1602
      243 -----SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
      1603 CACCTTCTGATCTCAGACCGAGACCGGACTTTTGGAAAGGAGACTGGACAGCCCACT 1662
      256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275
      1663 AAAAGCTTCGACCGCGAAACAGCTTGAGGCG----- 1695
      276 IleMetTyrSerAsnAspGluLeuProProGlnAspAspThrSerPheTyrSerGly 295
      1696 -----TACAGTCCG 1704
      296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHis 315
      1705 TCAGTGGCGACCGGAGGTGTAGCGTTGGGGGCGAGATGAGAACCTGTGTTGCCCAAG 1764
      315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
      1765 C-----CCTGGG-----AAGAGAGCGCGAAG 1785
      335 ngIylalagIuTyrAsnAspValalagIylusgluPro-----AsnTyrAsp 350
      1786 CGGAGAGAGCTTTCAGTGAACCTGGAGAGCAACCATTCACCATATGAGGAACGAGAT 1845
      350 gATgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----- 366

```

```

Db      1846 AGGACCAAGGAGATGGGAGGAGATCGAGCGGGGCTCCGACCGCACCCCTGAGCGCAG 1905
      367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg----- 378
      1906 CCGAAGAGAACCACTCCAGTAAAGGAGCAAGAGATCGACAGCAACTCCCTCAGCAAC 1965
      378 ----- 378
      1966 AGCAGCATGGGCTGAGGAAGGGGCCACACACACACACACAGAGCTGACACTCT 2025
      379 -----PrometArgSerAsp 384
      2026 TCCCAAGGAGAAAGGACAGACAGCGGCAATCACCGGACCAAGAGGCGCGAGCC 2085
      384 sAspSerTyrGluGlyTyrAspProAla-GlnLeuValArgLysSerLysArgThrArg 404
      2086 AAGCCTCTGAAAGAGCCAAACAGACAGACAAAGCTGAAGATTTTCAGAGTACGCT 2145
      404 rGAspGlySerValAspGlyPheThrProMetGlyValAspGluArg----- 419
      2146 CAGACACTACAGCTGGGTTGAAATGGGCTTCTGTGTGTGAAAAACAGCTGCTTCCCAAG 2205
      420 -----SerPheGlyArgGlySerValAlaAla----- 428
      2206 TCTATGATATATCTAGAGGGGAGCCAGGGGGTATCAGACAGTCTCTCAAGACACACT 2265
      429 -----ArgProIleArgGlyPro-ProAsp 436
      2266 TCTGGAGACAGCTGACCCAGCTGAAGATCGCCAGCGCTTCGACTGACAGCCAGCAAG 2325
      437 SerAspHisIleTyrArgGlyMetIleAlaLysGlyGlyThrPro----- 451
      2326 CTTCAGAGGTCAACAGAGCATC-----AAGCAGGGAGGCCCAACGGCTATGGCGTTC 2379
      452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
      2380 CTTTATAGCACCCAGCAACCCCGATGGGCTTGGACAGAG--CGATGCCACAGAT 2436
      472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
      2437 GAGCCCGTCTCAGAGCGGCGCTCTCAGAACTG----- 2472
      492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511
      2472 ----- 2472
      512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp----- 528
      2473 ---GTCCTCTACTTAAACAGAAAGCAGCGCGAGGGGTGATCAGCTTGCACAGTGGCGGG 2529
      529 -----AspGlyThr--ThrLeuPheLeuValProProSerAspPheLeuThr 543
      2530 TCCAAAGGAGAGAGAGCGACAGGCAATGCTTACGCTTCCACCCGTCGACATTTTCCAG 2589
      544 AspValLeuGln 547
      2590 CAGTACCTTCAG 2601

```

RESULT 13

ABL05395 standard; cDNA; 3084 BP.

ABL05395;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polymucleotide SEQ ID NO 10667.

Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ss.

Drosophila melanogaster.

XX

PN WC0200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB61292.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 10667; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB161676-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
 CC AB87072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SX Sequence 3084 BP; 706 A; 868 C; 885 G; 625 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,6e-06 Length: 3084
 Score: 287.50 Matches: 140
 Percent Similarity: 32.70% Conservative: 65
 Best Local Similarity: 22.33% Mismatches: 201
 Query Match: 5.99% Indels: 221
 DB: 4 Gaps: 27

US-09-920-705-3 (1-901) x ABL05395 (1-3084)

QY 11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
 Db 1111 GAGGATGATCCTCTTCTACGGCACACTGTTGCGAAGAACTCGAGGTTACCATCGCC 1170
 QY 31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyr 50
 Db 1171 GACGAGCAACTGCGCATCTTTGGCAATACGAGGTGATGACATGACATCAAG 1230
 QY 51 -----SerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluLeu 65
 Db 1231 CGACCCGCCCGGGTACGGGGAACGCGTTGCGGTGATCGTTACCAAACTCGATATG 1290
 QY 66 AlaValAlaAlaLysGlnLalaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
 Db 1291 GCTCACCGAGCGAAGATCGAATTGCTGACAGTACATTGGCAAGTTTCAGCGCAAGATC 1350
 QY 86 GluTyrAlaIleArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
 Db 1351 GGATCGGGGAAA---GTGACTCCGGCAACTCGATGATGATGATGAGGCTTAGTGCTCG 1407
 QY 106 ValSerLysAspAspLeuGlnGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
 Db 1408 ACTTCTGTACTCACTGAGCGAGAGGTTGATCTCTCGCGCATCAAGAAATTGAA 1467
 QY 126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGlyLysMetAspAspAlaLeuGln 145
 Db 1468 TACCAAGAGGGCGAGCCGTAAGCTTATATTCAGTACAGACAGTGGAGCGACCGCGCG 1527

QY 146 Ala---LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
 Db 1528 GCGTAAAGAGAGATCGAAGATTCCATTGGGTGACCGAGCGTCGATCGACGAC 1587
 QY 163 Phe-----LeuArgSerGlnAlaProLysLys 171
 Db 1588 TTGGCGAGTTGCCCGGCGCCACACCGGACGCGCTTTCAAAGAGTCTCAAGCGCCATAT 1647
 QY 172 GluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnIleLysProGlnTyr 191
 Db 1648 GATGAAGGCGGTGAGATAC-----AGACGTCGGGAATAC 1683
 QY 192 ProHisSerTyrGluAspPhe-----LysGlyAspValGlnProSerLys 206
 Db 1684 GATCCGATTACGAAGAGTCGCGCCCTACGCTCGCGGGGTGCTATTCG----- 1737
 QY 207 ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHis 226
 Db 1738 -----TATCACTAGGGGT----- 1752
 QY 227 AsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPhe 246
 Db 1753 -----GGCTACCGTGACGTGG--- 1770
 QY 247 AlaLeuValGluPheArgSerAlaGluGlnAlaArgGlnCysLysGluGlyLeuGlnGly 266
 Db 1771 -----GGATATCGCGGA 1782
 QY 267 ArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProGlu 286
 Db 1783 AGA-----GGAAAGGAGCATGATCATATATCAACAGACGTCGATAGGCCG--- 1827
 QY 287 GluAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn 306
 Db 1827 ----- 1827
 QY 307 AspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArg----- 324
 Db 1828 -----CCGATCTGCTGCTCTTAGGAGGTCTCTTCTTCAGTT 1866
 QY 325 ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValaGluTyrAsnAspValaGly 344
 Db 1867 CCCCCTCCGCGC-----GGA 1881
 QY 345 LysGluProAsnTrpArgArgProSerAlaAsnGly-----ThrGlyIleLeuProSer 362
 Db 1882 GTGAGATGATGATGCGACGCGCCGCGAGAAATCGTACGACAGAGAGCCGCTCCAGC 1941
 QY 363 ProThrGlyProGlyIle-----LeuProSerProAlaGlnGlyThrArgArg 378
 Db 1942 TCTCGGGAACCTGCTGTGAAGACGCTCTCGTTTGAATCCCATTTGAAGCGGCTCGCTCT 2001
 QY 379 Pro-----MetArgSerAsn----- 383
 Db 2002 CCCGATCTGATTCTGACACCTCGACTCGTGGAAACAGCCCTTGATCAGACGATACG 2061
 QY 384 ---ProAsp-----SerTrpGluGlyTyrAspProAlaGlnLeu 395
 Db 2062 GTTCCAGATGCGCGCGCAAGTCGACAGACAGTGTGACCGGA-----GGGCTAATC 2112
 QY 396 ValArgLysSer-----LysArgThrArgArgAspGlySerValaAspGlyPhe 411
 Db 2113 CTCAAGAGCTCACTGTTCCAGCGCAAGTTTCATTGACGAGCGAGATACGATATTGTG 2172
 QY 412 ThrProMetGlyValaAspGluArgSerPheGlyArgGlySerValaAlaAlaArgProIle 431
 Db 2173 GAGTCGCTGATGCGCGATGAGGAGGCGACACAACTATCATCAACCGACGATGCGGT 2232
 QY 432 ArgGlyProProAspSerAspHisIleTyrArgGlyMetIleAlaLysGlyGlyThrPro 451
 Db 2233 CTGATCTCGCCCAAGCTGCGACGTCGCAAAAGTATTCGATTCGATTCG----- 2283
 QY 452 ValCysAlaIleArgCysValaProMetGlyLysGlyIleGluThrLysLeuProGluVal 471

```

Db      2284 -----TGGACCGCATTTTCATGCTGTGGCAGGCTCCACA-----AACGAC 2325
QY      472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaIshisTyrAlaValAlaIle 491
Db      2326 ACCAATTGCG----- 2334
QY      492 GlyCyGluIleValPhePheValProAspArgGluGlnAspPheAlaSerTyrThrGlu 511
Db      2335 -----GACGATGCCAGCGCTGACAGCGCA 2358
QY      512 PheLeuArg-----TyrLeuSerSerIleAspArgIleGlyValAlaIleValLeu 527
Db      2359 CCACCTAAGAAACTTGGTTCCTTATCTTAACAGAGAGCGCGCTGCGTGAATCTGCTG 2418
QY      528 AspAspGlyThrThr-----LeuPheLeuValProProSerAspPheLeu 542
Db      2419 CTTAAAGAGAGACGAGGCACTGGGCTGCTGACGCTTCCCTCCCTCGACTTTTTC 2478
QY      543 ThrAspValLeuGlnValThr-----ArgGlnGluArg 553
Db      2479 ACAGAGCTGCTTAAAGCGCACTTGTACAGTGTGACGAGAGGCGCTGAAGAGATCAC 2538
QY      554 LeuTyrGlyValValLeuLys 560
Db      2539 CTGGTATTGTGTGTGCTGCC 2559

RESULT 14
ABL05394/c
ID      ABL05394 standard; cDNA; 9744 BP.
XX      ABL05394;
XX      26-MAR-2002 (first entry)
XX      Drosophila melanogaster expressed polynucleotide SEQ ID NO 10664.
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical; gene; ss.
XX      Drosophila melanogaster.
XX      WO200171042-A2.
XX      27-SEP-2001.
XX      23-MAR-2001; 2001WO-US009231.
XX      23-MAR-2000; 2000US-0191637P.
XX      11-JUL-2000; 2000US-00614150.
XX      (PEKE ) PE CORP NY.
XX      Venter JC, Adams M, Li PMD, Myers EW;
XX      WPI, 2001-656860/75.
XX      P-PsDB; ABB61291.
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
XX      PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX      PT interactions.
XX      Claim 1; SEQ ID NO 10664; 21PD + Sequence Listing; English.
XX      The invention relates to an isolated nucleic acid detection reagent
XX      CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX      CC useful in developmental biology and in elucidating cell signalling and
XX      CC cell-cell interactions in higher eukaryotes for the development of
XX      CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX      CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
XX      CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
XX      CC ABB72072). The sequence data for this patent did not form part of the
XX      CC printed specification, but was obtained in electronic format directly

```

```

CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX      SQ      Sequence 9744 BP; 2598 A; 2274 C; 2331 G; 2541 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.:      1,24e-05      Length:      9744
Score:      287.50      Matches:      140
Percent Similarity:      32.70%      Mismatch:      65
Best Local Similarity:      22.33%      Mismatch:      201
Query Match:      5.99%      Indels:      221
DB:      4      Gaps:      27
US-09-920-705-3 (1-901) X ABL05394 (1-9744)
QY      11 AspAspSerGlyPheGlnSerAsnAsnLeuTyrValGlySerLeuThrProGluThrThr 30
Db      7411 GAGATGATGCTCTTCTTCTACGCGCACACTGTTTGCAGAAATCTGAGGTTACATCGCC 7352
QY      31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyr 50
Db      7351 GACGACGAACGTGCTCGCATCTTTGCAATACGAGGTGTGATGATACATCGACATCAAG 7292
QY      51 -----SerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluGln 65
Db      7291 CGACCGCGCGCGGTACGGGGAACGGCTTCGGGTTTGTACGTTACCAAAACGTGATATG 7232
QY      66 AlaValAlaIleValGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleValIle 85
Db      7231 GCTCACCGAGCAAGATCGAATTCGTGACAGTACATTCGCAAGTTTCAGTGCAGAGATC 7172
QY      86 GluTyrAlaArgProAlaIleProCysIleSerLeuTyrValGlyIleGlyProAsn 105
Db      7171 GGATACGGGAAA--GTGACTCCGGCACTCGTATGTGATCGAGCTTAGGTGCTGG 7115
QY      106 ValSerIleAspAspLeuGluGlnIlePheSerIlePheGlyLysIleGluAspPheArg 125
Db      7114 ACTTCTGTCACTCAGCTGAGGAGAGGTTGATCGCTTGTGCGCATCAAGAAATTGA 7055
QY      126 PheLeuArgGluArgIleThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
Db      7054 TACCAAGAGGCGAGCGCTTATTCCTTATTCAGTACGAGACAGTGCAGCGCGCG 6995
QY      146 Ala--LysSerMetLysnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
Db      6994 GCGGTAAAGAGATGCGAGATTCCTATTGGTGAACCGAGCGTCGACGCGAC 6935
QY      163 Phe-----LeuArgSerGlnAlaProLys 171
Db      6934 TTTCGCGAATGCCCGCGCCACACCGGACGCGCTTCAAGACGTCAAGCCGCAAT 6875
QY      172 GluGlnTyrAlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyr 191
Db      6874 GATGAAGTGCCTGTGAGTAC-----AGAGTCCGGAATAC 6839
QY      192 ProHisSerTyrGluAspPhe-----LysGlyAspValGlnProSerLys 206
Db      6838 GATCCGATTACGAAGATCCCGCGCTACGCTCGCGGTGCTATTCTCCG----- 6785
QY      207 ValLeuTyrIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHis 226
Db      6784 -----TATCCACTAGGGGT----- 6770
QY      227 AsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPhe 246
Db      6769 -----GGCTACCGTGAAGTGGG--- 6752
QY      247 AlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlnIleLeuGlnGly 266
Db      6751 -----GATATTCGCGGA 6740
QY      267 ArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProGlu 286
Db      6739 AGA-----GGAAGAGCATGTACATTATCACAAGACGTGATAGCGG--- 6695

```

```

QY      287  GlnAspAspThrSerPheTyrSerGlyMetIleArgSerArgThrAspMetPheAsnAsn 306
DB      6695  -----
QY      307  AspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArg----- 324
DB      6694  -----
QY      325  ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValAsnIleTyrAsnAspValValGly 344
DB      6655  CCCCCCTCCGGC-----
QY      345  LysGluProAsnTyrPArgArgProSerAlaAsnGly-----ThrGlyIleLeuProSer 362
DB      6640  GTGAGAGATGATGGGCGGAGCGCCCGCAGAGATCGACAGAGAGAGAGCGCCGCTCCAGC 6581
QY      363  ProThrGlyProGlyIle-----LeuProSerProAlaGlnGlyThrArgArg 378
DB      6580  TCTGGGAACTGTGTGTAGAACGCTCTCGTTCTAGATCCCATTTGAAGCGGCTCGCTCT 6521
QY      379  Pro-----MetArgSerAsn----- 383
DB      6520  CCGGATCTGATTTCTGACACTGACTCGTGGAGACGAGCGCTTGATCAGCCAGTACG 6461
QY      384  ---ProAsp-----SerTyrGluGlyTyrAspProAlaGlnLeu 395
DB      6460  GTTCCAGATGTGGCGGCAAGTGCAGAGACAGTGTGACCGGA-----GGCTAATAC 6410
QY      396  ValArgGluSer-----LysArgThrArgArgAspGlySerValAspGlyPhe 411
DB      6409  CTCAAGAGCTCACTGTTCCAGCAGCAAGTTTATTCAGCAGCAGCATTCGATTTGTCG 6350
QY      412  ThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAlaIleAspProIle 431
DB      6349  GAGTCCCTATATGGCCATGAGAGGCGCAAGCACACCTTCGATTCACCCAGCGACTGGCT 6290
QY      432  ArgGlyProProAspSerAspHisIleTyrArgGlyMetIleAlaLysGlyGlyThrPro 451
DB      6289  CTGGATCCCGCCCAAGCTGAGCAGCATGCAAAAACGATTTGCAATCCCATCG----- 6239
QY      452  ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
DB      6238  -----TCGACGCCCATTTTCATGGGTCTCGCAGGGTCCACA-----AACGAC 6197
QY      472  ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysIleTyrAlaValAlaIle 491
DB      6196  ACCAATTGC----- 6188
QY      492  GlyCysGluIleValPhePheValProAspArgGluAspPheAlaSerTyrThrGlu 511
DB      6187  -----GACGATGCCAGCGCTGCAGACGGGA 6164
QY      512  PheLeuArg-----TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeu 527
DB      6163  CCACCTAAGAACTTGTTCTTCTATCTTAAAGCAGAAAGGAGCGGCTGTGATCTCGCTG 6104
QY      528  AspAspGlyThrThr-----LeuPheLeuValProProSerAspPheLeu 542
DB      6103  CTTTAAACAAGAGACGAGAGCAACTGGGGTGTGTACGCTTCCCTCCCTGCACCTTTTCC 6044
QY      543  ThrAspValLeuGlnValThr-----ArgGlnGluArg 553
DB      6043  AAGAGAGCTGCTAAGGCGACCTTGTCACAGTCTGACGAGAGAGGCGCTGAAGAGAGATCAC 5984
QY      554  LeuTyrGlyValValLeuLys 560
DB      5983  CTGGTATTTGTGTGTGTGCGC 5963

```

RESULT 15
 ABZ21697
 ID ABZ21697 standard; cDNA; 3383 BP.
 XX

```

AC      ABZ21697;
XX      27-FEB-2003 (first entry)
DT      Human RBM15-L protein encoding cDNA SEQ ID NO:11.
XX
DE      Human; RBM15; RNA binding motif protein 15; megakaryoblastic leukemia 1;
XX      MKL1; fusion protein; acute megakaryoblastic leukemia; AMKL; cytostatic;
XX      t(1; 22) chromosomal rearrangement; gene therapy; chromosome 1p13; gene;
XX      ss.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      CDS      84..2957
FT              /*tag=a
FT              /product="RBM15-L protein"
XX
PD      WO200288309-A2.
XX
PN      07-NOV-2002.
XX
PF      23-APR-2002; 2002WO-US012797.
XX
PR      27-APR-2001; 2001US-0286910P.
XX
PA      (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX      (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI      Morris SW, Ma Z, Hitzler JK;
XX
DR      WPI; 2003-103455/09.
XX
DR      P-PSDB; ABB56070.
XX
PT      New RNA-binding motif protein-15 (RBM15)-megakaryoblastic leukemia-1
XX      (MKL1), MKL1-RBM15-S and MKL1-RBM15-S+AR fusion proteins, useful for
XX      identifying agents useful for treating patients with acute
XX      megakaryoblastic leukemia.
XX
PS      Disclosure; Page 89-93; 109pp; English.
XX
CC      The present invention describes an RNA-binding motif protein-15 (RBM15)-
CC      megakaryoblastic leukemia-1 (MKL1) fusion protein, a MKL1-RBM15-S fusion
CC      protein, and a MKL1-RBM15-S+AR fusion protein associated with acute
CC      megakaryoblastic leukemia (AMKL). Also described: (1) an antibody that
CC      specifically binds to the RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR
CC      fusion proteins; (2) a non-human transgenic animal that has been altered
CC      to express a gene encoding a RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR
CC      fusion protein; (3) identifying an agent capable of binding to a RBM15-
CC      MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR fusion protein; (4) detecting the
CC      t(1;22) chromosomal rearrangement associated with AMKL; and (5) screening
CC      for agents capable of (selectively) inhibiting the activity of a fusion
CC      protein arising from the t(1;22) chromosomal rearrangement associated
CC      with AMKL. The fusion proteins have cytostatic activity and can be used
CC      in gene therapy. The RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR fusion
CC      proteins and nucleotide molecules are useful for designing and preparing
CC      agents that specifically inhibit the expression of the RBM15-MKL1 or MKL1-
CC      RBM15 genes in cells for therapeutic and other purposes. The transgenic
CC      animals are useful for identifying and testing carcinogenic or
CC      therapeutic compositions. The methods are also useful for detecting the
CC      t(1;22) chromosomal rearrangement associated with AMKL, or for
CC      identifying agents useful for treating patients with AMKL. The antibodies
CC      can be used to selectively kill cells expressing RBM15-MKL1, MKL1-RBM15-
CC      S, or MKL1-RBM15-S+AR fusion proteins. RBM15 is located to chromosome
CC      1p13, and MKL1 is located to chromosome 22q13. The present sequence
CC      encodes the human RBM15-L protein which is given in the exemplification
CC      of the present invention
XX
SO      Sequence 3383 BP; 812 A; 854 C; 952 G; 765 T; 0 U; 0 Other;

```

Alignment Scores: 1.53e-05 Length: 3383
 Pred. No.: 276.50 Matches: 162
 Score:

Percent similarity: 35.87%
 Best Local Similarity: 23.24%
 Query Match: 5.76%
 DB: 8
 Gaps: 34

US-09-920-705-3 (1-901) x ABZ21697 (1-3383)

QY ProPhArg-----AlaAspAspSerGlyPheGlnSerAenAanLeu 20
 DB CTTTCAAGAGAGAGTGGATGATTTTCAACCCGAGAGATGATCAGGAGCTAACCGGAGCTC 1211
 QY 1152 TTPValGlySerLeuThrProGluThrThrGlnSerAspLeuThrGluLeuPheGlyArg 40
 DB 1212 TTTCTGGGCAACTAGACATCAGCTGTAACGAGAGATTTTAAAGAGGGCTTGAATCGC 1271
 QY 41 TyrGlyAsp1LeaAspArg1LeuThrVal--TyrSerSerArgGly-----Phe 55
 DB 1272 TTTGGAGTCATCAAGAGATGATATCAGAGGCTTCTCGCGGCAAGACTAGACTATAC 1331
 QY 56 AlaPhe1LeuTyrThrArgHisValGluGluValAlaAlaValSerGluAlaLeuGlnGly 75
 DB 1332 GGCCTTCTCAAAATTGAGAACTTAGATATGCTCAACGGGCAAAATTAGCAATGCTGCG 1391
 QY 76 AlaAsnLeuAsnGlySerGln1LeuGly1LeuTyrAlaArgProAlaLysProCysLys 95
 DB 1392 AAAATTATATATGGGAATCTATCAAAATTGGTTATGCTAA--GCTACACCCACCACC 1448
 QY 96 SerLeuTTPValGlyGly1LeuGlyProAenValSerLysAspAspLeuGluGluPhe 115
 DB 1449 CGCCTCTGGGTGGAGGCTCGGACCTTGGGCTTCTCTGCTCGCTGCGACGAGAAATT 1508
 QY 116 SerLysPheGlyLys1LeuGluAspPheArgPheLeuArgLysThrAlaPhe1Le 135
 DB 1509 GATGATTTGGCAACATACCCACCATAGACTACCGAAAGGTATGTTGGCATATATC 1568
 QY 136 AspTyrTyrGluMetAspAspAlaLeuGlnAla--LysSerMetAsnGlyLysProMet 154
 DB 1569 CAGTAAAGAAAGCTGGATGACGAGCGCATGCTGCTGAGCCCATATGGGGCTTCCCAT 1628
 QY 155 GlyGlyLys-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGlu 172
 DB 1629 GGTGGCCCAAGATCAGCGCTTAGAGTAACTTTGCCGACACCAACATCCTTACCGAG 1688
 QY 173 GlnTyr-----AlaGlySerTyrAspAsn 180
 DB 1689 CAGTATCTGCAGCCTTCTGCTGACTATATGAGCTGGTGAAGATGCTTTTGGACAT 1748
 QY 181 Arg-----AsnGlyAsnMetAsn1LysProGlnTyrProHisSerTyr 195
 DB 1749 CGGGACACAGACCTTTGAGGGGTGCTCGGAGATAGACACCA-----CCCTTACTATAC 1802
 QY 196 GlnAspPheLysGlyAspValGlnProSerLysValLeuTyr1LeuGlyPheProPhe 215
 DB 1803 AGAGATCGATAGGACCTTATCTGACTGAT--TGGGTGCCACCCCAACCCCA 1859
 QY 216 AlaThrGlnCysAsnAspArgGln1LeuHisAsnAlaMet1LeuPheGlnGlu1Le 235
 DB 1860 GTCCGA-----GAACGACAGACTCGGACTGCACTTCTGTGCTGCTTAT 1907
 QY 236 GlnArgValLysSerTyrProSerArg-----AsnPheAlaLeuValGluPheArgSer 253
 DB 1908 GAGCCACTGGATAGCTCATGATCCGAGCGGAGATGCTGTGCTTGGACCGGACAGAGGT 1967
 QY 254 AlaGlnGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnProArg 273
 DB 1968 GATCGAGATCTGCCAGACAGAGAC-----CAGCTTAGG 2003
 QY 274 IleLys1LeuMetTyrSerAsnAspGluLeuProGluGlnAsp-----Asp 289
 DB 2004 AACGGA-----AGCTGCTGAGAGAGTGGAGAGACTCATCTGGAT 2045
 QY 290 ThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspProSer 309

DB 2046 AGGTCTCTGAGATGACCGCCACGAAACGT-----CAC 2081
 QY 310 CysValSerSerPro---HisSerThrGly1LeuProGlySerMetArgProLeuArgGly 328
 DB 2082 TGGCTCTCTTCTCTGACCGGAGTCCGAATTGAGAGATGAGCGGAGCTTACCAACACG 2141
 QY 329 ThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluProAen 348
 DB 2142 GACATATGATCGATCT-----TCCGCTTCTCTTGGAAAGCCCTCT 2183
 QY 349 TTPArgArgProSerAlaAsnGlyThrGly1LeuLeuProSerProThrGlyProGly1Le 368
 DB 2184 -----CCAATCAGAACGAGACGAGGTAGTTG----- 2210
 QY 369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyrGlu 388
 DB 2211 -----GAGAAAGCCAGGAGTACAGACGAGAC--CGTAAACCTTCATCATCGTGA 2261
 QY 389 GlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgAspArgLysSerVal 408
 DB 2262 -----CGAGATGAGAACACCGGACAACTGCTCCACT 2294
 QY 409 AspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGly-----Ser 425
 DB 2295 GAGGAAAAAGCCTCTGAAAAAGAACCCGCTGATGGAGTGCACCTAGACACACG 2354
 QY 426 ValAlaAlaArgPro1LeaArgGlyPro----- 434
 DB 2355 ACTGCTTCTCCAGGTGAAGTCCCTCCAGAAACAGATGGGGGGAAGGCCCTGTG 2414
 QY 435 -----ProAspSerAspHis1LeuTyrArgGlyMet----- 444
 DB 2415 GCATCAGCTCTCCAAACTCTGTTGGCTGCGAGGGATGCTTACTAGAAACACG 2474
 QY 445 -----IleAlaLysGlyGlyThrProValCysCysAlaArgCys 457
 DB 2475 AACTTCTCTCAACATGATGATGTTGAGGGTGAACCTCAAGTGGCTAGTATGCTTCT 2534
 QY 458 Val-----PrometGlyLysGly1LeuGlnThrLysLeuProGluValAlaAsnCys 474
 DB 2535 GTGGAGGTTCAACTGGAGGCAAGTGGCCAGCTCAAGATGATCACTCAGGCTCCGCTTG 2594
 QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAla-----IleGly 492
 DB 2595 GAC--CAGCCCAAGTGGATGAAGTAACTCAACGATCAAAATGACAGGGCCCAATGGT 2651
 QY 493 CysGlu1LeuAlaPhePheValProAspArgGluGlu----- 504
 DB 2652 TATGCCATCTTGTGGCTGTGCTGGAAGTCTGACAGCCGGTCCCTCTTCAGCT 2711
 QY 505 -----AspPheAlaSerTyrThrGlu-----PheLeuArgTyrLeuSer 517
 DB 2712 GATCAGACAGTCCGCACTTCTACTCAGAGGCCACTTAGGAACTGTGCTTATTAAG 2771
 QY 518 SerLysAspArgAlaGlyValAlaLysLysAspAspArgGlyThrThrLeuPheLeuValPro 537
 DB 2772 CAAGGACAGCAAGCGGGGTATGAGCTCCCTGTGGGGGCG----- 2813
 QY 538 ProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVal 557
 DB 2814 -----AACAAAGACAGGAAAAACCGCGGCTC 2840
 QY 558 ValLeuLysLeuProProAlaValAlaProValThrAlaSerTyrArgGlnGluSerGln 577
 DB 2841 CTTCAATGCTTCCCA----- 2855
 QY 578 SerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSer 597
 DB 2856 -----CCTTGGAGTCTCCAGACAGTTCCTGAAATCCCTGGCCAAAGGCA----- 2900
 QY 598 LeuTyrProProArgGluAsnTyr-----IleArgGlyAla 609
 DB 2901 CTGGCCAAATCTGAGAGATTAACCTGTCATGATCATTTGCTGGTGGC 2951

Tue Apr 12 08:34:21 2005

us-09-920-705-3.rng

Page 24

Search completed: April 8, 2005, 09:07:22
Job time : 1324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 08:37:33 ; Search time 370 Seconds

(without alignments)
3984.555 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801
Sequence: 1 MALSMKFRPADSGFQSNL.....LQIQKQKQSSGCTPAQGP 901

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlh
-Q=/cgm2_1/USPTO.epool_h/US09920705/runat_07042005_101814_20701/app_query.fasta_1.1095
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09920705.@CEN_1_1_75@runat_07042005_101814_20701 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgm2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgm2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgm2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgm2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgm2_6/prodata/1/ina/PCITUS_COMB.seq:*
6: /cgm2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271.5	5.7	11231	4	US-09-949-016-14030
2	268	5.6	3192	4	US-09-949-016-2288
3	268	5.6	3228	4	US-09-949-016-2288
4	246.5	5.1	1477	4	US-09-976-594-794
5	246	5.1	1228	1	US-07-726-607C-1
6	246	5.1	2238	1	US-07-843-949A-1
7	246	5.1	2238	2	US-08-218-978-1
8	246	5.1	2356	4	US-09-949-016-4865
9	246	5.1	2356	4	US-09-949-016-4865
10	242.5	5.1	1401	1	US-07-843-949A-3
11	242.5	5.1	1401	2	US-08-218-978-3
12	236.5	4.9	8147	3	US-09-514-247A-9

13	215.5	4.5	1968	4	US-09-949-016-3886	Sequence 3886, Ap
14	208	4.3	4674	4	US-09-410-551B-25	Sequence 26, Ap1
15	208	4.3	4674	4	US-09-940-316B-25	Sequence 26, Ap1
16	207	4.3	1323	4	US-09-949-016-4868	Sequence 4868, Ap
17	207	4.3	1323	4	US-09-949-016-4869	Sequence 4869, Ap
18	201.5	4.2	4737	4	US-09-410-551B-30	Sequence 30, Ap1
19	201.5	4.2	4737	4	US-09-940-316B-30	Sequence 30, Ap1
20	200.5	4.2	1467	1	US-07-881-075-50	Sequence 50, Ap1
21	200.5	4.2	1467	1	US-08-120-827-50	Sequence 50, Ap1
22	200.5	4.2	1467	1	US-08-478-675-50	Sequence 50, Ap1
23	200.5	4.2	1469	1	US-08-461-027-1	Sequence 1, Ap11
24	200.5	4.2	1469	3	US-09-082-501-1	Sequence 1, Ap11
25	200.5	4.2	1469	5	PCT-US92-01836-1	Sequence 1, Ap11
26	200.5	4.2	2559	4	US-09-920-132-135	Sequence 135, Ap
27	200.5	4.2	2562	1	US-08-146-421-4	Sequence 4, Ap11
28	200.5	4.2	13987	2	US-08-804-227C-13	Sequence 13, Ap1
29	200.5	4.2	44377	2	US-08-804-227C-7	Sequence 7, Ap11
30	200.5	4.2	44377	2	US-08-804-198-1	Sequence 1, Ap11
31	200	4.2	2642	4	US-09-949-016-1282	Sequence 1282, Ap
32	199.5	4.2	2731	4	US-09-976-594-281	Sequence 281, Ap
33	198.5	4.1	9046	1	US-08-227-536-1	Sequence 1, Ap11
34	198.5	4.1	9046	5	PCT-US95-04682-1	Sequence 1, Ap11
35	195.5	4.1	1231	4	US-09-949-016-1766	Sequence 1766, Ap
36	195.5	4.1	1416	3	US-08-866-928B-2	Sequence 2, Ap11
37	195.5	4.1	1416	4	US-09-685-836-2	Sequence 2, Ap11
38	195.5	4.1	4767	4	US-09-410-551B-28	Sequence 28, Ap1
39	195.5	4.1	4767	4	US-09-940-316B-28	Sequence 28, Ap1
40	195.5	4.1	4818	4	US-09-410-551B-32	Sequence 32, Ap1
41	195.5	4.1	4818	4	US-09-940-316B-32	Sequence 32, Ap1
42	194.5	4.1	2277	1	US-08-676-967-5	Sequence 5, Ap11
43	194.5	4.1	2277	1	US-08-676-974-5	Sequence 5, Ap11
44	194.5	4.1	2277	2	US-09-098-487-5	Sequence 5, Ap11
45	194	4.0	4725	4	US-09-410-551B-24	Sequence 24, Ap1

ALIGNMENTS

RESULT 1
US-09-949-016-14030 ; Sequence 14030, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for windows Version 4.0
SEQ ID NO 14030
LENGTH: 11231
TYPE: DNA
ORGANISM: Human
US-09-949-016-14030

Alignment Scores:
Pred. No.: 2.24e-11
Score: 271.50
Percent Similarity: 34.32%
Best local Similarity: 22.22%
Query Match: 5.66%
DB: 4
Gaps: 31

US-09-920-705-3 (1-901) x US-09-949-016-14030 (1-11231)
QY 7 ProPhenArg-----AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20

Db 3032 CTTTTCAGAGAGATGATGATTTCAACCGAGATATACAGACGATTAACCGAGCCTC 3091
Qy 21 TTPValGlySerLeuThrProGluThrThiGluSerAspLeuThrGluLeuPheGlyArg 40
Db 3092 TTCTTGGGCAACCTAGACATCATCTGTAAACGAGAGATTTTAAAGAGGCGTTGATCGC 3151
Qy 41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
Db 3152 TTTCGACTCATACAGAGATGATGATATCAAGAGCCTTCTCGCGGACGACTAGTACTTAC 3211
Qy 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlyArg 75
Db 3212 GCGCTTCTCAAAATTGAGAACTTAGATATGCTCACCGGCGCAAAATTAGCAATGCTCGC 3271
Qy 76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLys 95
Db 3272 AAAATTATTAATTCGAAATCTATCAAAATTGTTATGTAA---GCTACACCCACCACC 3328
Qy 96 SerLeuThrValGlyGlyIleGlyProAsnValSerLysAspLeuGluGluGluPhe 115
Db 3329 CGCCTCGGGTGGGAGGCTCGGACCTTGGCTTCTTGGCTCGCCTCGGACGAGAAATT 3388
Qy 116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGlyLysThrAlaPheIle 135
Db 3389 GATCGATTGTGGCACCATACGACCATAGACTACCGAAAGGTGATAGTGGGCATATATC 3448
Qy 136 AspTyrTyrGluLeuAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
Db 3449 CAATATGAAAGCCTGATGACGACGATGCTGCTGACACCATATGCGGGCTTCCACATT 3508
Qy 155 GlyGlyLys---SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGlu 172
Db 3509 GGTGGCCAGATCGACGCTTAGAGTAACTTTGCCGACACCGAAATGCTTACCGACGAG 3568
Qy 173 GlnTyr-----AlaGlySerTyrAspAsn 180
Db 3569 CAATATCTGACGCTCTGCTGCTGACTCATATAGACTGTGACAGATCTTTTGGACAT 3628
Qy 181 Arg-----AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
Db 3629 CGGGACACAGACCCCTTTAGAGGGTCTCGGAGATGAGACACA---CCCTTACTATAC 3682
Qy 196 GluAspPheLysGlyAspValGlnProSerLysValLeuTyrIleGlyPheProProThr 215
Db 3683 AGAGATCGTGAAGAGACCTTATCTGACTCTGAT---TGGGTGCCACCCCAACCCCA 3739
Qy 216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyLysIle 235
Db 3740 GTCCGA-----GACGCGACACTCGGACTGCACTTCTGTGCTGCTTAC 3787
Qy 236 GluArgValLysSerTyrProSerArgAsnPheAla----- 247
Db 3788 GAGCCACTGATAGCTAGATGCGACGGCGGATGTGTCTTGGACCGGACAGAGAGT 3847
Qy 248 -----LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlyLysLeu 264
Db 3848 GATCGAGATCTGCCAGCAGACAGAGACCGCTAGGAAGCGAAGCTGCTGAGAGAGT 3907
Qy 265 GlnGlyArgLeuPhe-----AsnAsnProArgIleLysIleMetTyr 278
Db 3908 GGAGAGCGTCATCTGGATAGTCTCTGAGAGTGAACCGCCAGCAAAAGCTACATGCGCT 3967
Qy 279 SerAsnAspGluLeuProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
Db 3968 CCTTCTCGTACCGGAGCTCAGATTTGAGC----- 3997
Qy 299 SerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGly 318
Db 3998 AGTACCGGAGATGTTACAACAGCAC-----AATGATCGATCTTCCCGT 4042
Qy 319 Ile-----ProGlySerMetArgProLeuArgGlyThrAsnGlnArgSerTyr 334
Db 319 Ile-----ProGlySerMetArgProLeuArgGlyThrAsnGlnArgSerTyr 334

Db 4043 CTTTCTTGGAAAGGCCCTTCCAAATCAGACACAGACAGAGTAGTTGGAGAGAGC--- 4099
Qy 335 AsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsnTyrArgProSerAla 354
Db 4099 ----- 4099
Qy 355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
Db 4100 -----CAG 4102
Qy 375 GlyThrArgArgProMetArgSerAsnProAspSerTyrGluGlyTyrAspProAlaGln 394
Db 4103 GGTGACACCGAGAC---CGTAAATCTGTCATGACTGAA----- 4141
Qy 395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
Db 4142 -----CGAGATAGAGAAAGACCGGACAACTGCTCCCACTGAGGAGAAAGCCCTCTG 4192
Qy 415 GlyValAspGluArgSerPheGlyArgGly-----SerValAlaAlaArgProIle 431
Db 4193 AAAAAAGAACCGCTGTGATGGAGTGACCTAGACACAGACATGCTTCTCCAAAGCTG 4252
Qy 432 ArgGlyPro-----ProAsp 436
Db 4253 AAGTCCCGTCCCAAGAACAGATGGGGGACAGCCCTGTGGCATCAGCTTCCCAA 4312
Qy 437 SerAspHisIleTyrArgGlyMet----- 444
Db 4313 CTGTGTTGGCTGGCAGAGGATGCTTCTACTGAAAGAACAGCAACTTCTTCCAAACATG 4372
Qy 445 ---IleAlaLysGlyGlyTyrProValCysValAlaArgCysVal-----ProMet 460
Db 4373 CATCTGTTGACGGGTGACCTCAAGTGGCTAGTGTCTTCTTGGAGAGGTTCAACTGGA 4432
Qy 461 GlyLysGlyIleGluThrLysLeuProGluValAlaAsnGlySerAlaArgThrAspLeu 480
Db 4433 GGCAAAGTGGCCAGCTCAAGATCATCAGCGTCTCGGTTGAGC---CAGCCCAAGTTG 4489
Qy 481 AsnMetLeuAlaLysHisTyrAlaValAla-----IleGlyCysGluIleValPhePhe 498
Db 4490 GATGAAGTAACTGACAGCATCAAGATGACAGGCGCCCAATGTATGCAATCTTTGGCT 4549
Qy 499 ValProAspArgGluGlu-----AspPheAlaSer 508
Db 4550 GTGCTGGAAGTTCTGACAGCGGCTCTCTCTTCTCAGCTGATAGACACTGCCACT 4609
Qy 509 TyrThrGlu-----PheLeuArgTyrLeuSerSerLysAspArgAlaGly 523
Db 4610 TCTACTCAGAGGCCACTTAGGAACCTTGTGTCTATTAAAGCAAAAGCAGGAGCGGG 4669
Qy 524 ValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThr 543
Db 4670 GTGATCAACCTCCTGTGGGGGAGC----- 4693
Qy 544 AspValIleuGlnValThrArgGlnGluArgLeuTyrGlyValIleuLysLeuProPro 563
Db 4694 -----AACAAAGACAAAGAAACCGGGTCTCTTACGTGCTTCCCA--- 4735
Qy 564 ProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyr 583
Db 4736 -----CTTGTGAGTTG 4747
Qy 584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProAlaGlu 603
Db 4748 TCCAGAGATTCCTGATTCCTCGGACGAGCA-----CTGGCCAAATCTGAAGAA 4798
Qy 604 AsnTyr-----IleArgGlyAla 609
Db 4799 GATTACTGTCATGATCATTCATTCCTGCTGCTG 4831
RESULT 2
US-09-949-016-2288
Sequence 2288, Application US/09949016

```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2288
; LENGTH: 3192
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2288

Alignment Scores:
Pred. No.: 6,07e-12 Length: 3192
Score: 268.00 Matches: 155
Percent Similarity: 34.28% Conservative: 86
Best Local Similarity: 22.05% Mismatches: 248
Query Match: 5.58% Indels: 214
DB: 4 Gaps: 30

US-09-920-705-3 (1-901) x US-09-949-016-2288 (1-3192)

QY 7 ProPhaArg-----AlaAspAspSerGlyPheGlnSerAsnLeu 20
DB 1032 CTTTACAGAAAGTGCATGATGATTTCACTCCAGAGATGATCAGAGCTAACCGAGCGCTC 1091
QY 21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
DB 1092 TTCTTGGCACTTACATCATCTGATGACGAGAGATTTTAAAGAGGGCGTTTGATCGC 1151
QY 41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
DB 1152 TTTCGAGTATCATCAGAGATGATATCAAGAGCGCTTCTCGCGCCAGACTAGTACTTAC 1211
QY 56 AlaPheIleTyrThrArgHisValGluGluAlaValAlaAlaGlyGluAlaLeuGlnGly 75
DB 1212 GCGCTTCTCAAAATTGAGAACTTGAATATGCTCACCGGGCCAAATTGAGCAATGCTGGC 1271
QY 76 AlaAsnLeuAsnGlySerGlnIleGlyIleGluTyrAlaArgProAlaLysProCysLys 95
DB 1272 AAAATTATATTTGGAAATCTTACAAATTTGTTATGTTAA--GCTACACCCACCACC 1328
QY 96 SerLeuTyrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluPhe 115
DB 1329 CGCCTCTGGGTGGAGGCGCTGGAGACTTGGGTTCTCTGCTGCGCGACAGAAATTT 1388
QY 116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
DB 1389 GATCGATTTGGCCACCTATCCACCATAGACTACCGAAAGTGATGTTGGCATATATATC 1448
QY 136 AspTyrTyrGluLeuSerAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
DB 1449 CAGTATGAAAGAGCTGGATGCGAGCATGCTGCTGGACCCATATGGGGGCTTCCCACTT 1508
QY 155 GlyGlyIle---SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGlu 172
DB 1509 GGTGGCCAGATGAGAGCGCTTAGAGTAGACTTTGGCGACACGAAATGCTTACAGAGCAG 1568
QY 173 GlnTrp-----AlaGlySerTyrAspAsn 180
DB 1569 CAGTATCTGACAGCGCTGCTGACCTTGAATCATATAGAGCTGTGTGACAGATGCTTTGGA 1628
QY 181 Arg-----AsnGlyAsnMetAsnIleLysArgProGlnTyrProHisSerTyr 195

```

```

DB 1629 CCGGCACACAGACCTTTGAGGGGTGCTCGGATAGGACACCA-----CCCTACTATATAC 1682
QY 196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProThr 215
DB 1683 AGAGATCGTATGAGGACCTTTATCTGACTGTAT---TGGGTGACACCCCAACCCCA 1739
QY 216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
DB 1740 GTCCGA-----GAGCGCAGACTCGGACTGACACTTCTGTGCTGCTTAC 1787
QY 236 GluArgValLysSerTyrProSerArgAsnPheAla----- 247
DB 1788 GAGCCACTGATAGACTTAATGACGAGCGGATGTTGCTTGGACCGGACAGAGGT 1847
QY 248 -----LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluLysLeu 264
DB 1848 GATCGAGATCTGCCACAGACAGACAGACAGCTTAGAGAAAGCTGCTGAGAGAGAT 1907
QY 265 GlnGlyArgLeuPhe-----AsnAsnProArgIleLysIleMetTyr 278
DB 1908 GAGAGACGTATCTGATAGTCTCTGAGAGTACCGCCCAAGAAAGCTACCTGCGCT 1967
QY 279 SerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
DB 1968 CTTCTCCCTGACCGCGACATCCAGATTTGAC----- 1997
QY 299 SerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGly 318
DB 1998 AGTAGCGGGATCGTTACAAACAGCGAC-----AATGATCATCTTCCCT 2042
QY 319 Ile-----ProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyr 334
DB 2043 CTTCTTGAAAGCGCTCTCCATCATCAGACAGACAGAGTATGTTGAGAAAGAC--- 2099
QY 335 AsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsnTrpArgArgProSerAla 354
DB 2099 ----- 2099
QY 355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
DB 2100 -----CAG 2102
QY 375 GlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGln 394
DB 2103 GGTACACACGAGAC---GTAATAAATCTGATCAGCTGAA----- 2141
QY 395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
DB 2142 -----CGAGATAGAAAGACACCGACAACTGCTCCACATGAGGAAAGAAACCTCTG 2192
QY 415 GlyValAspGluArgSerPheGlyArgGly-----SerValAlaAlaArgProIle 431
DB 2193 AAAAAGAAAGACCGCTGTATGAGATGACACTAGACAGACAGCACTCTTCCCAAGCTG 2252
QY 432 ArgGlyPro-----ProAsp 436
DB 2253 AAGTCCCGCTCCAGAAACAGATGGGGGAGACGCCCTGTGGCATCAGCTTCCCAAA 2312
QY 437 SerAspHisIleTrpArgGlyMet----- 444
DB 2313 CTCTGTTTGGCTGCGAGGAGCATTCCTTCTTACATGAAAGACAACTTCTTCCCAAGT 2372
QY 445 ---IleAlaLysGlyGlyThrProValCysValAlaArgCysVal-----ProMet 460
DB 2373 CATCTGTTCAGAGGTACTCCAAAGTGGCTAGATGCTTCTTGGAGGGTTCAACTGGA 2432
QY 461 GlyLysGlyIleGluThrLysLeuProGluValAlaAsnCysSerAlaArgThrAspLeu 480
DB 2433 GGCAGAGTGGCCAGCTCAAGATCACTCAGCGCTCCGTTTGGAC---CAGCCCAAGTTG 2489
QY 481 AsnMetLeuAlaLysHisThrAlaValAla-----IleGlyCysGluIleValPhePhe 498

```

```

Db      2490 GATGAGTAAGTCTGACGATCAAGTAAGGAGGCCAATGTTATGCATTTCTTTGGCT 2549
Qy      499 ValProAspArgGluGlu-----AspPheAlaSer 508
Db      2550 GTGGCTGGAACTTGTACACGCGGTCTCTCTTCTCAGTGCATGCAGACACGCGCCT 2609
Qy      509 TyrThrGlu-----PheLeuArgTyrIleuSerSerIleuAspArgAlaGly 523
Db      2610 TCTACTCAGAGGCCACTTAGAACCTTGTGTCTATTAAAGCAAAAGCGAGCGCGG 2669
Qy      524 ValAlaIleuAspArgGlyThrThrLeuPheLeuValProPheSerPheLeuThr 543
Db      2670 GTGATCAGCCTCCTGTGGGGGCGC----- 2693
Qy      544 AspValLeuGluValThrArgGluGluArgLeuTyrGlyValIleuLeuIleuProPro 563
Db      2694 -----AACAAAGACAGAGAAACACCGGGGTCTTCATGCTTCCCA--- 2735
Qy      564 ProAlaValProValThrAlaSerTyrArgGluGluSerGlnSerGlnSerAsnProLeuHisTyr 583
Db      2736 -----CCTTGTGAGTTC 2747
Qy      584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu 603
Db      2748 TCCGACGAGTTCTGTGATTCCTGCGCAAGCA-----CTGGCCAATCTGAAGAA 2798
Qy      604 AsnTyrIle 606
Db      2799 GATTACTG 2807

RESULT 3
US-09-799-451-752
Sequence 752, Application US/09799451
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, ping
APPLICANT: Goodrich, Ryle
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhang, Jie
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Ma, Yungling
APPLICANT: Yamazaki, Victoria
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Yonghong
APPLICANT: Wehrman, Tom
APPLICANT: Ghosh, Reena
APPLICANT: Dimanac, Radoje T.
TITLE OF INVENTION: No. 6783969el Nucleic Acids and
FILE REFERENCE: 803
CURRENT APPLICATION NUMBER: US/09/799,451
CURRENT FILING DATE: 2001-03-05
NUMBER OF SEQ ID NOS: 948
SOFTWARE: pt_FL_genes Version 2.0
SEQ ID NO 752
LENGTH: 3228
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (108)..(2882)
US-09-799-451-752

Alignment Scores:
Pred. No.: 6,18e-12 Length: 3228
Score: 268.00 Matches: 155
Percent Similarity: 34.28% Conservative: 86

```

```

Best Local Similarity: 22.05% Mismatches: 248
Query Match: 5.58% Indels: 214
DB: 4 Gaps: 30

US-09-920-705-3 (1-901) x US-09-799-451-752 (1-3228)

Qy      7 ProPheArg-----AlaAspAspSerGlyPheGlnSerAsnLeu 20
Db      1044 CTTTCAGAGAGATGATGATGATTTACCCCGAGATGATCAGCGATTAACCGAGCTC 1103
Qy      21 TrpValGlySerLeuThrProGluThrGluSerAspLeuThrGluLeuPheGlyArg 40
Db      1104 TTTCTGGCAACCTTACATCACTGTAACGGAGATGATTAAGAAAGCGCTTGATGCG 1163
Qy      41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
Db      1164 TTTGAGATCATCAGAGAGATGATTAAGAGAGCGCTTCCCGGCACAGTAACTTAC 1223
Qy      56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaIleArgValAlaLeuGlnGly 75
Db      1224 GGCCTTTCAAATTTGAGAACTTAAATGATGATGCTCACCGGCGCAATTAGCAATGCTGGC 1283
Qy      76 AlaAsnLeuAsnGlySerGlnIleGlyIleGluTyrAlaArgProAlaIleProCysIle 95
Db      1284 AAATTTAATTCGGAATCTTCAAAATTTGTTATGTA---GCAACCCACACACC 1340
Qy      96 SerLeuTrpValGlyGlyIleGlyProAsnValSerIleAspAspLeuGluGluPhe 115
Db      1341 CGCCTCTGGTGGAGGCGCTGGACCTTGGGCTTCCTTGGCTGCGCCAGAGAAATTT 1400
Qy      116 SerIlePheGlyIleGluAspPheArgPheLeuArgIleGluArgIleAlaPheIle 135
Db      1401 GATCATTTGGACACCATTCATCGACATTAAGCTTAAGAGGAGTAACTTGAGCTTAATC 1460
Qy      136 AspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
Db      1461 CAGTATGAAGACCTTGATGACCGCATGCGCTGCGACCCATAGCGGGGCTTCCACTT 1520
Qy      155 GlyGlyIle-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysIleGlu 172
Db      1521 GGTGGCCAGATGACCGCTTAAAGTAACTTTCGACACCGAACATCTTACACAGCAG 1580
Qy      173 GlnTrp-----AlaGlySerTyrAspAsn 180
Db      1581 CAGTATCGACGCTCGCTTACCTTATGATGAGCTGTGACAGATGCTTTGACAT 1640
Qy      181 Arg-----AsnGlyAsnMetAsnHisLysAspProGlnTyrProHisSerTyr 195
Db      1641 CGGGCACACAGACCTTTGAGGGGTGCTGGGATGAGACCA-----CCCTTAATATAC 1694
Qy      196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProPheThr 215
Db      1695 AGAGATCGTGAATGAGGACCTTATCTGACTGTAT---TGGGTGCGACCCACCCCA 1751
Qy      216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyIle 235
Db      1752 GTCCGA-----GAACGAGACATCGAGTGCACCTACTTGTGCTGCTTAC 1799
Qy      236 GluArgValLysSerTyrProSerArgAsnPheAla----- 247
Db      1800 GAGCACTGATGATGCTTAATTCGACGCGGATGCTTGTGACCGGACAGAGGT 1859
Qy      248 -----LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGluLeu 264
Db      1860 GATCGAATCTGCCACAGACAGACAGACACCTTAGAGAGAGAGGCTGCTGGAGAGT 1919
Qy      265 GlnGlyArgLeuPhe-----AsnAsnProArgIleLysIleMetTyr 278
Db      1920 GAGAGACGTATCATGTATAGTCTCTGAGAGTGAACCGCCACAGAAAGTACTCGCGT 1979
Qy      279 SerAsnAspGluLeuProProGluGlnAspPheThrSerPheTyrSerGlyMetLysArg 298
Db      1980 CTTTCTCGTACCGCAGTCCAGAAATTGAC----- 2009

```

```

QY 299 SerArgThrAspMetPheIleuAsnAspProSerCysValSerSerProHisSerThrGly 318
DB 2010 AGTAGCCGGAGTCGTTACACACAGCGAC-----AATAGTCGATCTTCCTCCGT 2054
QY 319 Ile-----ProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyr 334
DB 2055 CTTCCTTGGAAGAGCCCTCTCCAAATCAGAGACAGACAGAGTGTGGAGAAAGAC--- 2111
QY 335 AsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgProSerAla 354
DB 2111 ----- 2111
QY 355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
DB 2112 -----CAG 2114
QY 375 GlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGln 394
DB 2115 GGTGACACAGGAGAC---CGTAAAACTCTGCATCAGCTGA----- 2153
QY 395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
DB 2154 -----CGAGATAGGAAGACCGGACCAACTGCTCCACAGGAGAAAAGCCCTCTG 2204
QY 415 GlyValAspGluArgSerPheGlyArgGly-----SerValAlaAlaArgProIle 431
DB 2205 AAAAAAGAACCGCTCTGATGGAGTGACACTAGACACGACTCTTCCTCCAAAGCTG 2264
QY 432 ArgGlyPro-----ProArg 436
DB 2265 AAGTCCCTCCGACGAAACAGATGGGGGAGACGCCCTGTGCATCAGCTCTCCCAA 2324
QY 437 SerAspHisIleTrpArgGlyMet----- 444
DB 2325 CTCTGTTGGCTGCGACGAGGCGATCTTCTACTGAAGACAGCACTTCTCTCCAAAGT 2384
QY 445 ----IleAlaLysGlyGlyThrProValCysCysAlaArgCysVal-----Promet 460
DB 2385 CATCTGTTCAGAGGTACCTCCAAAGTGGCTAGTAGTCTTCTTGAAGGTTCACTGGA 2444
QY 461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCySerAlaArgThrAspLeu 480
DB 2445 GCGAAAGTGGCCCACTCAAGATCACTCAAGCTCTCCGTTGGAC---CAGCCCAAGTTG 2501
QY 481 AsnMetLeuAlaLysHisTyrAlaValAla-----IleGlyCysGluIleValPhePhe 498
DB 2502 GATGAAGTAACTGACGACATCAAAAGTAGAGAGGCCCAAGTGTATGCAATCTTTGGCT 2561
QY 499 ValProAspArgGluGlu-----AspPheAlaSer 508
DB 2562 GTGCCCTGGAAGTTCTGACAGCCGCGTCTCTCTCTCAAGCTCATCAGACCTGCCACT 2621
QY 509 TyrThrGlu-----PheLeuArgTyrLeuSerSerLysAspArgAlaGly 523
DB 2622 TCTACTCAGAGGCCCACTTAGGAACCTTGTCTCTATTTAAGCAAAAGCAGGACCGGG 2681
QY 524 ValAlaLysLeuAspArgGlyThrThrLeuPheLeuValProProSerAspPheLeuThr 543
DB 2682 GTGATCAGCTCCCTCGGGGGG----- 2705
QY 544 AspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProPro 563
DB 2706 -----AACCAAGCAAGAAACACCGGGGCTCTCATGCTTCCCA--- 2747
QY 564 ProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyr 583
DB 2748 -----CCTGTGAGTTT 2759
QY 584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu 603
DB 2760 TCCACACAGATTCTCGATTCCCTCGCAAGCA-----CTGGCAAAATCGAAGAA 2810

```

```

QY 604 AsnTyrIle 606
DB 2811 GATTACTGT 2819

RESULT 4
US-09-976-594-794
; Sequence 794, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Farness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 794
; LENGTH: 1477
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4436028CB1
US-09-976-594-794

Alignment Scores:
Score: 1,09e-10 Length: 1477
Percent Similarity: 39.53% Matches: 99
Best Local Similarity: 23.29% Mismatches: 172
Query Match: 5.13% Indels: 85
DB: 4 Gaps: 21

US-09-920-705-3 (1-901) x US-09-976-594-794 (1-1477)
QY 4 SerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGly 23
DB 74 TCGTTCGACACCCACCATGATGAGAGCAGCGG---CAGCCCGGACCTCTATACGTAGGT 130
QY 24 SerLeuThrProGluThrLysLeuSerAspLeuThrGluLeuPheGlyArgTyrGly--- 42
DB 131 AACCTTCCAGAGATGTGACAAAGTCTTATCTTCACTGATGCTCAATGGAGCC 190
QY 43 -----AspIleAspArgIleThrValTyrSerSerArgGly---PheAlaPheIle--- 58
DB 191 TGTAAAGCTGTAAATATGATTAACAGCATACAGCAATGACCAATATGCTTGTGGAA 250
QY 59 TyrTyrArgHisIleValGluValAlaValAlaLysGluValLeuGlnGlyAlaAsnLeu 78
DB 251 TTTATGAAACAC---AGAGATGACGCTGCTGATTAAGCTATGATGAGGAGAAATTT 307
QY 79 AsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCysLysSer--- 96
DB 308 TTGGAAAGAGGCTCAAGTAACTGAGCAACCAACCAAGTAGCCAGAAAAGTAGTACT 367
QY 97 -----LeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeu 111
DB 368 TCCAACTACTTCATGATGTTGTGGGATTTGAATGCCAAGAAATTACACAGAAATATC 427
QY 112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLys 131
DB 428 AAATCAGATTGCCCCCTTGTGTAATAATGCGATGCCCGGTAGTTAAAGCATGCGCA 487
QY 132 Thr-----AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
DB 488 ACTGAAAATCCAAAGCTATAGCTTTTATCTTTTATAACAACTGATGCAAAAT 547
QY 146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
DB 548 GCGATTGTGCATATGCGCGGTCAAGTGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 607

```

QY 165 -----ArgserGlnAlaProLysLeuGluGlnTrpAlaGlySerTyrAspAsnArgAsn 162
Db 608 ACTGTTAAACCACTGGACCTTAAAGTACACAA-----GAAACCAACACT 652
QY 183 GlysMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
Db 653 AAGCAGTTGAGA-----TTGAAAGTGTAGTAAACCACTCA 668
QY 203 GlnProSerLys-----ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db 689 AGTCACAAAATGTTACTGTGTACTGTGAGAGAAATTCGCTGGGTAAACA----- 729
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGluIleGluValArgValLysSer 240
Db 740 ---GATCAGCTTATGAGACACAGACTTCTCAACCTTTGACAAATTTATGAAATTAAGAGTT 796
QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db 797 TTCCAGAAAAGGGGTATTCATTTGTCAGATTTTCAAGATTTCAAGATGACAGCCCATGGC 856
QY 261 LysGluGluLysLeuGlnArgLysLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db 857 ATTGTTTCGGTGAACGGTACTACAGATTGAAGACATGTGGTTAAA---TGCTATTGGGGT 913
QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db 914 AAGAATCTCCT-----GATATGACTTAAACTTCAACAGGTGACTATGACTCA 964
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyLeuPro 320
Db 965 TGG----- 967
QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
Db 968 GGCATATGAGCCCAAGTATGGA-----AACCCACAACAGTATGGA 1009
QY 341 AspValValGlyLysGluProAsnTrpArgArgPro----- 352
Db 1010 CACTATATGTCAAAT-----GGGTGGCAAGTACCGCTTATGAGATATACGGGCAACA 1063
QY 353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro-----GlyIle 368
Db 1064 TGAATACAAACAGAGATTGGAGTATGATCAATCACTTCTGCTGTGATGGATGGATTT 1123
QY 369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
Db 1124 GGTGTCAGCTCCCAAGGACAGCTCCTCCCTGTAATACCTCTCTTAACCAAGCC 1183
QY 389 GlyTyrAspProAla 393
Db 1184 GGATATGTATGGCA 1198

RESULT 5
US-07-726-607C-1
Sequence 1, Application US/07726607C
Patent No. 5298407
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
APPLICANT: Streuli, Michel
APPLICANT: Schloegman, Stuart F.
TITLE OF INVENTION: DNA ENCODING A PROTEIN
TITLE OF INVENTION: ACTIVE IN LYMPHOCYTE-
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/726,607C
FILING DATE: July 10, 1991
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2228
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-726-607C-1
Alignment Scores:
Pred. No.: 2,286-10 Length: 2228
Score: 246.00 Matches: 85
Percent Similarity: 39.65% Conservative: 74
Best Local Similarity: 21.20% Mismatches: 158
Query Match: 5,128 Indels: 84
DB: 1 Gaps: 17
US-09-920-705-3 (1-901) x US-07-726-607C-1 (1-2228)
QY 20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db 81 CTATACGTCGGTAAACCTTTCCAGAGATGTGACAGAACTCTTAATTCGAACTCTTTAGC 140
QY 40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly-----Phe 55
Db 141 CAGATTGACCTGTGAAATACTGCAAAATGATATGATACAGCTGAATGATCCCTAT 200
QY 56 AlaPheIleTyrTyrArgHisValGluGluAlaAlaAlaLysGluAlaLeuGlnGly 75
Db 201 TGTTTGTGAGTTTCATGAGCATGTCATGACAGCTGCACATTAAGCTGATGATGGA 260
QY 76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
Db 261 CCGAAGATATAGGTGTAAGAGATGAAATGGGCAACACCTTACAGCTCAAAAG 320
QY 95 LysSer-----LeuTrpValGlyLysIleGlyProAsnValSerLys 108
Db 321 AAAGATACAGCAATCATTTCCATGCTTTGTGTGATCTCAGCCCAAGAAATTACACT 380
QY 109 AspAspLeuGluGlnGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db 381 GAAGATATTAAGAGCTGCTTTGACCATTTGGAAGATATCAATGCCAGGTGTAAMA 440
QY 129 GluArgLysThr-----AlaPheIleAspTyrTyrGluMetAspAsp 142
Db 441 GACATGCAACAGGAAGTCTAAGGATATAGCTTGTCTCTTTTCAACAAATGGAT 500
QY 143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db 501 GCTGAAAACCCCATTTCAACAGATGGGTGCCAGTGGCTGTGTGAAGACAAATCAAGACT 560
QY 162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181
Db 561 AACTGGGCAACCGAAGAGCTCCCGCTCCAAAG-----AGTACATATGATGATCAAT 611

```

QY 182 AsnGlyAsnMetAsnHisIleuProGlnTyrProHisSerTyrGlnAspPheGlyAsp 201
DB 612 ACCAAACAGCTA-----TCATATGATGAGGTTGTAATCAG 647
QY 202 ValGlnProSerIleu---ValIleuTyrPheGlyPheProProThrAlaThrGlnCysAsn 220
DB 648 TCTAGTCCAAAGCACTGACTGATGACTGGA-----GGTGTACTCTTGGGCTA 698
QY 221 AspGlnGlnIleuHisAsnAlaMetIleuPheGlyGlnIleuGlnValIleuSer 240
DB 699 ACAGAAACAACTAATGCGTCAGACTTTTCCACCAATTGGCAAAATGAATGAATTCGAGTC 758
QY 241 TyrProSerArgAsnPheAlaLeuValGlnPheArgSerAlaGlnIleuValArgGlnCys 260
DB 759 TTTCAGATTAAGATATTCATTGTTGTCGTTCAATTCATGAAGTGCAGCATGCA 818
QY 261 LysGlnGlnIleuGlnIleuGlnIleuPheAsnProArgIleuIleuMetTyrSerAsn 280
DB 819 ATTGTTCTGTATGATGACTGACCATGAGTCATGTCATGTAATGCTATTGGGCGCAA 878
QY 281 AspGlnLeuProGlnIleuAspPheThrSerPheTyrSerGlyMetLysArgSerArg 300
DB 879 GAAACTCTT-----887
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
DB 888 ---GATATGATTAATCC-----GTGCAACAGCAGAAATCAATGGATATCC 932
QY 320 -----320
DB 933 CAACCTTATGGCCAGTGGGCGCAGTGTATGAAATGCACAAATTTGGCCAGTATATG 992
QY 321 -----GlySerMetArgProLeuArgGlyThrAsnGlnArgSerTyrAsnGlnIleGln 338
DB 993 CCTAATGTTGGCAGATTCCTGCATATGATGATGATGCGCATGGAACCGACAGCA 1052
QY 339 TyrAsnAspValIleGlyLysGlnProAsnTyrArgArgProSerAlaAsnGlnIleGln 358
DB 1053 TTTAATCAGACACAGTTCTTGCACCA---TGATGGGACCAAAAT-----TATGCA 1100
QY 359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlnIleGln 376
DB 1101 GTGCAACCG---CCTCAAGGGCAAAATGCGACGATGTTGCCCAATGACCTTGGGTAT 1157
QY 377 Arg 377
DB 1158 CGA 1160

```

RESULT 6

```

US-07-843-949A-1
Sequence 1, Application US/07843949A
Patent No. 5340935
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
APPLICANT: Streuli, Michel
APPLICANT: Schlosser, Stuart F.
TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/843,949A

```

```

FILING DATE: 19920219
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2228
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-843-949A-1

```

Alignment Scores:

Pred. No.:	2,28e-10	Length:	2228
Score:	246.00	Matches:	85
Percent Similarity:	39.65%	Conservative:	74
Best Local Similarity:	21.20%	Mismatches:	158
Query Match:	5.12%	Indels:	84
DB:	1	Gaps:	17

US-09-920-705-3 (1-901) x US-07-843-949A-1 (1-2228)

```

QY 20 LeuTyrValGlySerLeuThrProGlnTyrThrGlnSerAspLeuThrGlnLeuPheGly 39
DB 81 CTATACGTGGTAACTCTTCCAGAGATGACAGAGTCAATCTGCAACTTTAGC 140
QY 40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgIle-----Phe 55
DB 141 CAGATTGACCTTTGAAAACTGCAAATGATTGATGATGACCTGAAATGATCCCTAT 200
QY 56 AlaPheIleTyrTyrArgHisValGlnIleAlaValAlaLysGlnIleGlnIleGln 75
DB 201 TGTGTTGTGAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 260
QY 76 AlaAsnLeuAsnGlySerGlnIleLysIleGlnTyrAla---ArgProAlaLysProCys 94
DB 261 CGGAGATATATGGGTAAAGAGTCAAGTGAATGATGGGCAAAACCCCTAGCACTCAAAAG 320
QY 95 LysSer-----LeuTyrValGlyGlyIleGlyProAsnValSerLys 108
DB 321 AAAGATACAGCAATCATTTCCATGCTTTGTTGGTGAATCTCGCCCAAAATTAACAAT 380
QY 109 AspAspLeuGlnGlnIlePheSerLysPheGlyLysIleGlnAspPheArgPheLeuArg 128
DB 381 GAAGATATTAAGCTCTTTGACACATTTGAAAGAAATATCAGATCCCGAGTGTAAAA 440
QY 129 GlnArgLysThr-----AlaPheIleAspTyrTyrGlnLysAspArg 142
DB 441 GACATGCAACAGAAAGTCTAAGGATATGCTTTGCTCTCTTTTCAACAAATGAGAT 500
QY 143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuAlaGln 161
DB 501 GCTGAAAGCCCATTTAAACAGATGGGTGGCGAGTGGCTTGTGGAAGACAAATCAACAAT 560
QY 162 AspPheLeuArgSerGlnAlaProLysLysGlnGlnTyrAlaGlySerTyrAspAsnArg 181
DB 561 AACTGGGCAACCGAAAGCTCCCGTCCAAAG-----AGTACATATGATGATCAAT 611
QY 182 AsnGlyAsnMetAsnHisIleuProGlnTyrProHisSerTyrGlnAspPheGlyAsp 201
DB 612 ACCAAACAGCTA-----TCATATGATGAGGTTGTAATCAG 647

```

QY 202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db 648 TCTAGTCCAAACCACTGACTGATATCTGCA-----GGTGTACTTCTGGGCTA 698
QY 221 AspGluGlnIleLeuHISaenAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db 699 ACAAGAACCAACTAAAGCCGTCAGACTTTTTCACCATTTGGACAAATTAATGCAAAATTCGAGTC 758
QY 241 TyrProSerArgAsnAlaLeuValGluPheArgSerAlaGluAlaArgGlnCys 260
Db 759 TTTCAGATTAAGATATTCATTGTTGCGTCAATTCCTCAAGTGAAGTCAGACATGCA 818
QY 261 LysGluGlnLeuGlnGlyArgLeuPheAsnProArgIleLysIleMetLysSerAsn 280
Db 819 ATTGTTCTGTTATGATGACTACCATTTGAAGTCATGTTGAAATGCTTATGGGGCAA 878
QY 281 AspGluLeuProProGluGlnAspPheThrSerPheTyrSerGlyMetLysArgSerArg 300
Db 879 GAACCTCTT----- 887
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHISerThrGlyIlePro 320
Db 888 ---GATATGATTAATATCCC-----GTGCACACAGCAATCAATGATGATATCCC 932
QY 320 ----- 320
Db 933 CAACCTTATGCGCAGTGGCGCAGTGGTATGAAATGCAACAAATTTGGCCAGTATATG 992
QY 321 -----GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
Db 993 CCAATATGTTGGCAAGTCTCGCATATGAAATGATATGATGCGCAGGACATGGAACCAAGCA 1052
QY 339 TyrAsnAspValValGlyLysGluProAsnTyrArgArgProSerAlaAsnGlyThrGly 358
Db 1053 TTTAATCAGACACAGCTTCTGCACCA---TGGATGGACCAAT-----TATGCA 1100
QY 359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlnGlyThr 376
Db 1101 GTTCAACCG---CCTCAAGGGCAAAATGCGACATGTTGGCCCAATCAGCTTCTGGGTAT 1157
QY 377 Arg 377
Db 1158 CGA 1160
RESULT 7
US-08-218-978-1
; Sequence 1, Application US/08218978
; Patent No. 5837811
; GENERAL INFORMATION:
; APPLICANT: Anderson, Paul J.
; APPLICANT: Streufl, Michel
; APPLICANT: Schlossman, Stuart F.
; TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
; NUMBER OF SEQUENCES: 9
; LYMPOCYTE-MEDIATED CYTOTOXICITY
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218,978
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/843,949

FILING DATE: February 19, 1992
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2228
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-218-978-1
Alignment Scores:
Pred. No.: 2,28e-10 Length: 2228
Score: 246.00 Matches: 85
Percent Similarity: 39.65% Conservative: 74
Best Local Similarity: 21.20% Mismatches: 158
Query Match: 5.12% Indels: 84
Gaps: 17
US-09-920-705-3 (1-901) x US-08-218-978-1 (1-2228)
QY 20 LeuTrpValIleGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db 81 CTATGCTGCTGATTAATCTTCCAGAGATGTGACAGAGCTTAATTCGCAACTCTTTAGC 140
QY 40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly-----Phe 55
Db 141 CAGATTGACCTTGAATAAATCGCAAAATGATTTATGATATGAGTGAATGATATGATATG 200
QY 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db 201 TGTTTTGTGAGATTTCATGACATGCTCATGACCTGACCTGACATTAAGCTGATATGAATGA 260
QY 76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---AspProAlaLysProCys 94
Db 261 CGAAGATTAATGCGTGAAGAGTCAAGATGATGGCAACACCCCTTACAGATCAAAAG 320
QY 95 LysSer-----LeuTrpValGlyGlyIleGlyProAsnValSerLys 108
Db 321 AAAGATACAAAGCATCATTTCCATGCTCTTGTGTTGATGATCGACCCAGAAATTAACAAT 380
QY 109 AspAspLeuGluGlnGluPheSerLysPheGlyLysIleGluPheAlaPheLeuArg 128
Db 381 GAAGATTAATAAGCTGCTTTTGACCACTTTGGAAGATATCAAGATGCCAGTGGTAAA 440
QY 129 GluArgLysThr-----AlaPheIleAspTyrTyrGluMetAspAsp 142
Db 441 GACATGGCAACAGAAAGTCAAGGATATGCTTTCTCTCTTTTCAACAAATGGAT 500
QY 143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyLysSerPheLeuArgVal 161
Db 501 GCTGAATAACCGCATTCACAGATGGTGGCCAGTGGCTTGTAAGAACAAATCAAGAACT 560
QY 162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaIleLysTyrAspAsnArg 181
Db 561 AACTGGGCAACCGAAAGCCCTCCGCTCCAAAG-----AGTCAATATGATGATCAAT 611
QY 182 AsnGlyAsnMetAsnHisLysProGluTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db 612 ACCAAACAGCTA-----TCATATGATAGATGTTGAATTCAG 647
QY 202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220


```

Db      1122 CCTAATGTTGGCAAGTTCCTGCAATGGAATGATGCGCAGCATGGAACCGAAGGA 1181
Qy      339 TyrAsnAspValValGlyLysGluProAsnTPrArgRProSerAlaAsnGlyThrGly 358
Db      1182 TTTAATCAAGACAGACTCTTCGACCA---TGATGGAGACCAAT-----TATGGA 1229
Qy      359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlyThr 376
Db      1230 GTGCACCG---CCTCAAGGGCAAAATGAGCAGCATGTGGCCCAATCAGCCTTCGGGTAT 1286
Qy      377 Arg 377
Db      1287 CGA 1289

RESULT 9
US-09-949-016-4866
; Sequence 4866, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4866
; LENGTH: 2356
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4866

Alignment Scores:
Pred. No.:      2,486-10      Length:      2356
Score:          246.00      Matches:      85
Percent Similarity: 39.65%      Conservative: 74
Best Local Similarity: 21.20%      Mismatches: 158
Query Match:      5.12%      Indels:      84
DB:               4          Gaps:      17

US-09-920-705-3 (1-901) x US-09-949-016-4866 (1-2356)
Qy      20 LeuTPrValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db      210 CTATACGTCGGTACCTTTCCAGAGATGAGCAGAAAGCTCTAATTCGTGCACTCTTTAGC 269
Qy      40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly-----Phe 55
Db      270 CAGATTGACCTTGTAATAAACTGCAAAATGATTATGATACAGCTGGAATGATCCCTAT 329
Qy      56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaIleGlyLalaLeuGlnGly 75
Db      330 TGTTTTGTCGAGTTTCATGACATGTCATGCACTGCGCACTTACGCTATGATGATGGA 389
Qy      76 AlaAsnLeuAsnGlySerGlnIleLeuGlyIleGlyTyrAla---ArgProAlaIleProCys 94
Db      390 CCGAAGATATATGCGTAAAGAGTCAAGTGAATGGAGCAACAACCCCTTAGCAGTCAAAAG 449
Qy      95 LysSer-----LeuTPrValGlyGlyIleGlyProAsnValSerIys 108
Db      450 AAAGTTACAGCAATCATTTCCATGTCCTTGTGTGATCTCAGCCGCAAAATTAACAAC 509
Qy      109 AspAspLeuGluGluIlePheSerIysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db      510 GAAGATATATAAAACGCTTTTGCACACATTTGGAAGAAATATACAGATGCCGAGGTATAAA 569

```

```

Qy      129 GluArgLysThr-----AlaPheIleAspTyrTyrGluMetAsp 142
Db      570 GACATGCGAAGCAAGAAAGCTTAAGGATATATGCTTGTCTCTTTTCAACAAATGCGAT 629
Qy      143 AlaLeuGlnAla---LysSerMetAsnGlyLysPrometGlyIleSerPheLeuArgVal 161
Db      630 GCTGAAMACGCCATTCAACAGATGGGTGCGCAGTGGCTGTGGTGAAGACAATACAGAACT 689
Qy      162 AspPheLeuArgSerGlnAlaProLysGluGlnTPrAlaGlySerTyrAspAsnArg 181
Db      690 AACTGGGCAACCCGAAAGCCTTCCTCCCAAG-----AGTACATATGAGTCAAT 740
Qy      182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db      741 ACCAAACAGCTA-----TCATATGATGAGTGTAAATCAG 776
Qy      202 ValGlnProSerLys---ValLeuTPrIleGlyPheProProThrAlaThrGlnCysAsn 220
Db      777 TCTAAGTCCAAAGCACTGACTGATATACGTGGA-----GGTGTACTTCTGGGCTA 827
Qy      221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db      828 ACAGAAACAATCAATGCGTCAGACTTTTCCACATTTGGACAAATATATGAAATTCGATC 887
Qy      241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db      888 TTTCCAGATATAAGATATTCATTGTTGGTTCAATTCCTCAGTAAAGTCAGCAGCATGCA 947
Qy      261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db      948 ATTGTTCTGTATATGTAATGTAACCAATGCAAGTCATGTTGAAATGCTATTTGGGCGAAA 1007
Qy      281 AspGluLeuProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db      1008 GAAACTCT----- 1016
Qy      301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
Db      1017 --GATATGATATAATCC--GTGCACAGCAGCAATCAAAATGATATATCCC 1061
Qy      320 ----- 320
Db      1062 CAACCTATATGCGCAGTGGGCGCAGTGGTATGGAATCAACAATAATGGCCAGTATATG 1121
Qy      321 -----GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
Db      1122 CCTAATGTTGGCAAGTTCCTGCAATGGAATGATGATGCGCAGGATGGAACGACAGGA 1181
Qy      339 TyrAsnAspValValGlyLysGluProAsnTPrArgRProSerAlaAsnGlyThrGly 358
Db      1182 TTTAATCAAGACAGCTCTTCGACCA---TGATGGAGACCAAT-----TATGGA 1229
Qy      359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlyThr 376
Db      1230 GTGCACCG---CCTCAAGGGCAAAATGAGCAGCATGTGGCCCAATCAGCCTTCGGGTAT 1286
Qy      377 Arg 377
Db      1287 CGA 1289

RESULT 10
US-07-843-949A-3
; Sequence 3, Application US/07843949A
; Patent No. 5340935
; GENERAL INFORMATION:
; APPLICANT: Anderson, Paul J.
; APPLICANT: Streuli, Michel
; APPLICANT: Schloesman, Stuart F.
; TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

```

```

ADDRESSER: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/843,949A
FILING DATE: 19920219
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1401
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-843-949A-3

Alignment Scores:
Pred. No.: 2 14e-10 length: 1401
Score: 242.50 Matches: 99
Percent Similarity: 39.29% Conservative: 68
Best Local Similarity: 23.29% Mismatches: 173
Query Match: 5.05% Indels: 85
DB: 1 Gaps: 21
US-09-920-705-3 (1-901) x US-07-843-949A-3 (1-1401)
QY 4 SerMetLysProPheArgAlaAspSerGlyPheGlnSerAsnAsnLeuTrpValGly 23
DB 31 TCGGTGGACCCGACCATGATGGAAGACGAGG---CAGCCCGGACTCTATACGTAGT 87
QY 24 SerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgGly--- 42
DB 88 AACCTTCCAGAGATGACAGAGAGCTCTTATCTTCAGTGTTCAGTCAGATTGACCC 147
QY 43 -----AspIleAspArgIleThrValTyrSerSerArgGly---PheAlaPheIle--- 58
DB 148 TGTAAAGCTGTAAATGATACAGACATACAGACATGACCCATATGCTTTGGAA 207
QY 59 TTTTATGAACAC---AGAGATGACGCTGCTGATTCGCTTGTGATGGAGAAAAATT 264
DB 208 TTTTATGAACAC---AGAGATGACGCTGCTGATTCGCTTGTGATGGAGAAAAATT 264
QY 79 AsnGlySerGlnIleLeuIleGlyTyrAla---ArgProAlaLysProCysLysSer--- 96
DB 265 TTGGGAAAGAGGTCAAGTAACTGGGCAACACACACAGTACCGAGAAAAAGATTACT 324
QY 97 -----LeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeu 111
DB 325 TCCAAATCATTCCATGCTGTTGTTGGGATTGATGCTCCAAATTCACACAGAGATATC 384
QY 112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLys 131
DB 385 AATTCAGCATTTCCCTTGGTAAATATCGAGTCCCGGGTATGTTAAAGACATGGCA 444

```

```

QY 132 Thr-----AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
DB 445 ACTGGAATAATCCAAAGCTATGTTGTTATCTTTTATTAACAACTGATGCAAGAAAT 504
QY 146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
DB 505 GCCATTGTGCATATGGCGGTGAGTGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 564
QY 165 -----ArgSerGlnAlaProLysLysGluGlnIleGlySerTyrAspAsnArgAsn 182
DB 565 ACTGCTAAACCAACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 609
QY 183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
DB 610 AGCAGCTTGAAGA-----TTGTAAGATGTTAGTAAACCACTCA 645
QY 203 GlnProSerLys-----ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
DB 646 AGTCCAAAATAATTTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCT 696
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
DB 697 ---GATCAGCTATATGACAGACAGATTCACCATTTGGACAAATTTGGAATAAGAGTT 753
QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
DB 754 TTGCCAGAAAAGGGCTATTCATTTGTCAGATTTTCAACCATGAAAGTCCAGCCCATGCC 813
QY 261 LysGluGluIleGluGlnIleValGluPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
DB 814 ATTGTTTCGCTGACGCTACTACGATTTGAAGACATGTCGTTTAA---TGCTATTGGGCT 870
QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
DB 871 AAAAGATCTCTCT-----GATATGACTTAAAACTTCCACAGAGTGAATGATGATCA 921
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
DB 922 TGG----- 924
QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValGluTyrAsn 340
DB 925 GGCCTAATGAGCCCAAGCTGATGGA-----AACCCACAAACGATATGCA 966
QY 341 AspValValGlyLysGluProAsnTrpArgArgPro----- 352
DB 967 CAGTATATGCAAT-----GGTGGCAAGTACCGCTTATGAGATATACGGCAACCA 1020
QY 353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro-----GlyIle 368
DB 1021 TGGAAATCAACAAAGATTTGGAGATGATCAATACCTTCGCTGCTGATGGATGGATTT 1080
QY 369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
DB 1081 GGTGCTCAGCTTCCCAAGAGCAAGCTCTCCCTGTTATATCTCTCTTAACCAAGCC 1140
QY 389 GlyTyrAspProAla 393
DB 1141 GGATATGATATGCA 1155

```

```

RESULT 11
US-08-218-978-3
; Sequence 3, Application US/08218978
; Patent No. 5837811
; GENERAL INFORMATION:
; APPLICANT: Anderson, Paul J.
; APPLICANT: Streuli, Michel
; APPLICANT: Schloesman, Stuart F.
; TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
; LYMPOCYTE-MEDIATED CYTOTOXICITY
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

```

```

ADDRESS: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218.978
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/843,949
FILING DATE: February 19, 1992
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1401
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-218-978-3

Alignment Scores:
Pred. No.: 2.14e-10 Length: 1401
Score: 242.50 Matches: 99
Percent Similarity: 39.29% Conservative: 68
Best Local Similarity: 23.29% Mismatches: 173
Query Match: 5.05% Indels: 85
Gaps: 21
US-09-920-705-3 (1-901) x US-08-218-978-3 (1-1401)

QY 4 SerMetLysProPhehArgAlaAspSerGlyPheGlnSerAsnLeuTyrValGly 23
DB 31 TCGGTGCAACCCGACATGATGAGAGCGG--CAGCCCGGACTTATACGTAGGT 87
QY 24 SerLeuThrProGluThrGlnuSerAspLeuThrGluLeuPheGlyArgTyrGly 42
DB 88 AACCTTCCAGAGATGTGACAGAGCTTATACCTTCACTTTCAGTCAGATGGACCC 147
QY 43 -----AspIleAspArgIleThrValTyrSerSerArgGly--PheAlaPheIle 58
DB 148 TGTAAAGCTGTAAATGATTAACAGAGCATACAGCAATGACCCATATGCTTGTGGA 207
QY 59 TTYTYTARGHIEValGluGluAlaValAlaIleValGluAlaLeuGlnGlyAlaAsnLeu 78
DB 208 TTTTATGAACAC--AGAGATGACGTGCTGCTATTAAGCTGTATGAATGGAGAAAATT 264
QY 79 AsnGlySerGlnIleLeuIleGluTyrAla--ArgProAlaIleProCysIleSer 96
DB 265 TGGGAAAGAGAGGTAAAGTAAACGTGGGACCAACCAAGTAAAGCAAAAAGATATCT 324
QY 97 -----LeuTyrValGlyGlyIleGlyProAsnValSerIleAspLeu 111
DB 335 TCCAAATCACTTCAGATGCTTGTGTGGGATTTGAAGTCCAGAAATTAACAAGAGATATC 384
QY 112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgIleArgLys 131

```

```

DB 385 AAATCAGCATTTGCCCTTTGTGTAATATATGATGCCCGGGTATGTAAGACATGGCA 444
QY 132 Thr-----AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
DB 445 ACTGGAATAATCCAAAGGCTATGTTTGTATCTTTTATTAACAACGTAGCAAAAAT 504
QY 146 Ala-----LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
DB 505 GCGATTGTCATATGCGCGCTCAGTGGTGGTCTGCTCAAAATCCGAACCAATGGGCC 564
QY 165 -----ArgSerGlnAlaProLysLysGluIntPrAlaGlySerTyrAspAsnArgAsn 182
DB 565 ACTCGTAAACCACTGCACCTTAAGTACACA-----GAAAACCAACT 609
QY 183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
DB 610 AAGCAGTTGAGA-----TTTGAAGATGTAGTAAACCACTCA 645
QY 203 GlnProSerLys-----ValLeuTyrIleGlyPheProPheThrAlaThrGlnCysAsn 220
DB 646 AGTCCAAAATAATTGACTGTGTGACTGTGAGAAATGCGTGGGTTAAC----- 696
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
DB 697 ---GATCAGCTTATGAGACAGACATTCACACATTTGGACAAATTATGAAATAAGATT 753
QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
DB 754 TTGCCAGAAAGAGGGTATTCATTTGTCAGATTTTCAACCCATGAAGTCCAGCCATGCC 813
QY 261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
DB 814 ATTGTTGGTGAACCGTACTACATTCAGAGACATGTGTTTAA---TGTATTGGGGT 870
QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
DB 871 AAAGATCTCT-----GATATGACTTAAACCTTCCAAACGTTGACTATATGTCAA 921
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
DB 922 TGG----- 924
QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
DB 925 GGCCAATGAGACCAAGTATGA-----AACCAACAACGTATGA 966
QY 341 AspValValGlyLysGluProAsnTyrArgArgPro----- 352
DB 967 CAGTATATGGCAAT-----GGGTGCAAGTACCGCTTATGAGATACGGGCAACCA 1020
QY 353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrArgPro-----GlyIle 368
DB 1021 TGGAAATCAACAAGATTTGGATGATCAATCACTTCTCTGCTGGATGGGTGATTT 1080
QY 369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyrGlu 388
DB 1081 GGTGCTCAGCCTCCCAAGGCAAGCTCCTCCCTGTATACCTCTCTTAACAAGCC 1140
QY 389 GlyTyrAspProAla 393
DB 1141 GGATATGTATGGCA 1155

RESULT 12
US-09-514-247A-9
Sequence 9, Application US/09514247A
Patent No. 6365361
GENERAL INFORMATION:
APPLICANT: TANABE SEIYAKU CO. LTD.
APPLICANT: TANIGUCHI, Tomoyasu
APPLICANT: MIZUKAMI, Junko
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA
FILE REFERENCE: TANIGUCHI=6

```

[illegible]

Db	5824	CCCTGCCAGAGACAGACAGACTGGGAGATTCTCTCTTCCGCGCGCTCCAAAGTGTCAACGC	5883
Qy	418	GluaSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProPro-----	435
Db	5884	TCTCATCATCTGGTGGAGCTGCACACCCAGGCGAGAACCCCTTGTCTTACACCTGCACAG	5943
Qy	436	AspSerAspHisIleTTPArgGlyMetIleAlaValGlyGlyThrProValCysCys----	454
Db	5944	AGTGCAGACACCAACGTGGAGA-----CGCGTGCACATGCACGTGTGTCCAGG	5991
Qy	455	-----AlaArgCysValPromet-----	460
Db	5992	ACTACGACCTTCGATCAACTGCTCTTAAACAGGAAGCGCATGCCCATTAAGATGGTGAAGT	6051
Qy	461	---GlyTyrGlyIleGlyThrLysLeuProGluValValencySerAlaArgThrAsp	479
Db	6052	GGGGGCTTGCGCTGGATGACGAGGCGAGACGCCAGGCGAGACCATGTCAAAGAGACCCC	6111
Qy	480	LeuAsnMetLeu-AlaLysHisTyrAlaVal-----AlaIleGlyCysGluIleValPhe	497
Db	6112	AGGAGTCAACCGCGCTGAGCATCCAGCGGTGCATCCAGTCGTGTGACCGCGTGCAC--	6165
Qy	497	ePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSe	517
Db	6170	---GTGCGCGACACG-----CACTGCTCGTGCATCT	6201
Qy	517	rSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrIlePheLeuValPr	537
Db	6202	GCCAGAAAGATAAACGGGTGTGTGCACACCAAGG-----	6238
Qy	537	oProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVal	557
Db	6239	-----CTGCAAAACCAAGACCAACCAAGCGGGCGTCCGCGGTGT	6273
Qy	557	1-----ValLeuLysLeuProProProAlaValProValThrAlaSerTyr	572
Db	6274	GCAAGCAGCTCATGCGCTCTGTCTGTCTTACACCGCAAGCATGCGCAAGAAACAATGCC	6333
Qy	572	rArg-----GlnGluSerGlnSerAsnProLeuHisTyrTyrMetAspGlnAl	587
Db	6334	CCGTGCGCTTGTGCTCAACATCAACACAAAGCTCCGCGACGACAGATCCAGACACGCGC	6393
Qy	587	AspAspSerProAlaAsnAla-----SerHisSerLeuTyrProProArgGluAs	604
Db	6394	TGCAGCAGAGCCCAACCTCATGCGCGCGAGATGCGACACATGMAACCGGCAACCTGCTCTC	6453
Qy	604	nTyrIleArgGlyLysProGluHisLeuThrAlaAlaSerLysProSerValSerGluPr	624
Db	6454	A---GCAAGTCTCTCTTCTCTTACCTTCAGACCGCGCGGAGCCCCACACAGACGCCCA	6510
Qy	624	oLeuAlaGlyLeuProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLe	644
Db	6511	G---CACACCCCAACCGCGGAGCCCTTGCAGCCGCCCAACCTTCAACCGCTAG----	6562
Qy	644	uAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro-----	660
Db	6563	---CATGTCAACGACTGTGCTTCCCAAGCGGTGCCGAGATCTACAGCCCCCAACAGCGGT	6618
Qy	661	-----GluSerIleGlnPrometSerGlyProSerThrValValSerThrAlaHisGlu	678
Db	6619	CCACAGAGAAACCTTACAGCAGCGGTGCGGCGCCGCCAC-----	6658
Qy	678	nSerAsnGlyLeuTyrAsnGlyGluAlaProSer--GlnAlaTPrLysArgGlyProGln	697
Db	6659	-----CCGCGCCAGCGCCCTCTCTGACGCGGTGGAAGCGGCT--CGGACG	6701
Qy	698	ThrValHisAspAlaSerAsnGlnSer-----PheGlnGlnTyrGlyAsnGlnTyrThr	715
Db	6702	ATTCAGAGGTGAGCCAGCAGACAGACACCTGTACCGGAGAAATCAATCAACAGCATGT	6761
Qy	716	ProAlaGlyGln-----LeuProProProProSerArgTyrTyrProProAlaSerAsnAsn	733
Db	6762	CCCCCAGACGACGAGGATGGGAGCCCGGAGGACAGATGGCCCTCTGTGAGCTGAAT	6821

```

QY 734 -----ProaenYrThrSerGlyMetValHisGlyAsnMet---GlnYrGlnSer 749
DB 6222 GTGCCCCGACCCCAACGAGTGAAGCGGCGCCGTCAATGCCAGTCTCCCGGAGATGG 6881
QY 750 GlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnYrSer 769
DB 6882 CAGCAGCGCGCCCTTCCCAAGCAGCAGCCCATGCGCTTGGCCAGCCTGTGATATCC 6941
QY 770 MetYrThrGlnGlySerSerAsnHisPro----- 779
DB 6942 ATGAGGCCCCAGCGCGCTGAGCTGCGCCCGATGCCAGCGCCAGCCAGAGAG 7001
QY 780 -----ValSerGlnPromet 784
DB 7002 ATCTACCCAGCGCTCTGCAAGACCTGCTGGAGACCTGAATGCGCCAGCTCTCCCTCAG 7061
QY 785 ValGlnGln----- 787
DB 7062 CAGCAACAGCAGGTGTGAACATTCTCAATCAACCCGAGCTAATGGCAGCTTTCATC 7121
QY 788 -----TyrGlnProGlnAlaSerMet 794
DB 7122 AAACAGCGCAACGCAAGTACGTGGCCATCAACCGCGCATGCAAGCCCGCTGAGCTTC 7181
QY 795 ProAsnGln---AsnYrGlyProIleProSer---TyrGlnGlnAlaAsnPheHisGly 812
DB 7182 CAGTCCAGCGCCGCGCATGCAACCCAGCCTGGCATGACACAGCCCGCAGCCTG 7235
QY 813 ValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAsnMetGln--- 831
DB 7236 -----CAGAACCTGAAAT-----GCCATGACGGCT 7259
QY 832 -----ProProAlaAspGlyAla---AsnLeuGlnProGln 842
DB 7260 GCGGTGCGCGCGCGCGCGCTGCTCCAGCAGCAGCAGCGAGCGCTGAAACCCCGAG 7319
QY 843 AsnGlnAlaLeuArgGlnGlnPrometIleSerGlyAspGlyGlnGlyThrThrAspGly 862
DB 7320 GCGCAGCGCTTGAACATCATG-----AACCCAGCAGCAACCCCAAC 7361
QY 863 GluValAspGlyAsnGlnArgGlyGlnSerThrLeuGlnPheAlaAsnLeuLeu 882
DB 7362 ATGGCAGATATGATCCACAGTACCGAAGATGTTACG-----AGCAGCTGCTGCAG 7415
QY 883 GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 900
DB 7416 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 7469

RESULT 13
US-09-949-016-3886
/ Sequence 3886, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 3886
/ LENGTH: 1968
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-3886

```

```

Alignment Scores:
Pred. No.: 6,22e-08 Length: 1968
Score: 215.50 Matches: 101
Percent Similarity: 36.95% Conservative: 59
Best Local Similarity: 23.33% Mismatch: 178
Query Match: 4,49% Indels: 96
DB: 4 Gaps: 17

```

US-09-920-705-3 (1-901) x US-09-949-016-3886 (1-1968)

```

QY 11 AAPAPSerGlyPheGlnSerAsnLeuPyrValGlySerLeuThrProGlnThrThr 30
DB 291 GAGCAGAC-----AAGACCACTCATGCACTCACTGCCCCAGAACATGACC 341
QY 31 GlnSerAspLeuThrGlnLeuPheGlyArgYrGlyAspIle----- 44
DB 342 CAGATGAGATTCAAGATCTCTTGGCAGCATGGCGCATCGAGTCTCGCAAGTTGTT 401
QY 45 ---AspArgIleThrValYrSerSerArgGlyPheAlaPheIleYrYrArgHisVal 63
DB 402 CCGGACAAGATCACA---GGCGAGCCTTGCTACAGGATTTGTGAACTATTCTGACCCC 458
QY 64 GluGlnAlaValAlaAlaGlyGlnAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 83
DB 459 ATATGATGACAGCAAGCCATCAACACCTCAACGCGCTCAATTACAGACAGAACATC 518
QY 84 LysIleGluYrAlaArgProAlaLeuProCys-----LysSerLeuTrpValGly 100
DB 519 AAGGTGCTTATGCGACAGCCAGTTCAGATTCATCCGGATGCTAATCCTGTACGTGAC 578
QY 101 GlyIleGlyProAsnValSerLysAspAspLeuGlnGlnGlnPheSerLysPheGlyLys 120
DB 579 GGGCTCCCAAGACCATAGCCAGAAAGATGAGAGCCTTCTTCCAGTACGCGCCG 638
QY 121 IleGluAspPheArgPheLeuArgGlnArgYrThrAlaPhe-----IleAspYr 137
DB 639 ATCATCAGTCCCGCATCTGTGTGACAGGTCAACAGTCTCTCGGGGTGTGGATTC 698
QY 138 TyrGlnMetAspAspAlaLeuGlnAla-----LysSerMetAsnGly---Lys 152
DB 699 ATCCGCTTTGACAAAGAGATTGAGCGCGGAGAGCTTCAAGAGATGGAATGGCAGAG 758
QY 153 PrometGlyLysSerPhe---LeuArgValAspPheLeuAsnSerGlnAlaProValLys 171
DB 759 CCGCTGGCGCAGCTGAGCCCATCAAGTTCGGAACCAACCCAGCAGAGACG 818
QY 172 GluGln-----TrpAlaGlySerYrAspAsnArgAsnGlyAsnMet 185
DB 819 GGGCAGGCGCTGCTACCCACCTCTACCAATCATCCCGCGCTACGACAGGCCCTTA 876
QY 186 AsnHisLysProGln----- 190
DB 879 CACCATCAGACCGCGCTTCCGGCTGAGACATTGCTCAACATGCGCTACGCGTCAAG 938
QY 191 ---TyrProHisSerYrGluAspPheGly-----AspVal 202
DB 939 AGCTTTCGCGCATCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 998
QY 203 GlnProSerLysValLeuTrpIleGlyPheProThrAlaThrGlnGlnGlnGlnGln 222
DB 999 GCGCGCGCGCGCGCGCGCTGCTGATCTGTGTAACAACCTGTACCGAGGCGAGAG 1058
QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGlnIleGlnArgValLys----- 239
DB 1059 AGCGTGTGTGGCAGCTGTTCGGGCTTTGGGCGACATCCCAACGTCAGATCATCCGT 1118
QY 240 ---SerYrProSerArgAsnPheAlaLeuValGlnPheArgSerAlaGlnGln 256
DB 1119 GATTTCACACCAACAGTGAAGGTTTGGCTGTGTGACATGACCACTATGACAG 1178
QY 257 AlaArgGlnCysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 276

```

Db 1179 GCGGCATGCGCATCGCAGCCTGAACGCTATCGCTGGGAGCGGTGTCAGATC 1238
Qy 277 MetTyrSerAsnAspGluLeuProProGluInAspAspThrSerPheTyrSerGlyMet 236
Db 1239 TCCTTC----- 1244
Qy 297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSer 316
Db 1245 AATACACGAAACAGCAGCAGGCGGTGAGCCACCCCGCTGCC-TCACACCC----- 1297
Qy 317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
Db 1298 -----CTCCCGGCGGCGAGAGAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAG 1351
Qy 337 AlaGluTyrAsnAspValValGlyLeuGluProAsnTrpArgArgProSerAlaAsnGly 356
Db 1352 GCCCAAGAGAGAG-----CAAGAGCAGCCCGCAGCAGAGAGAGAGAGAGAGAGAG 1390
Qy 357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376
Db 1391 -----GAGGCGCCACGCTCCTGCGGAAGCA---CAGGTTGAG 1426
Qy 377 ArgArgProMetArgSerAsnProAspSerTrpGlnGly 389
Db 1427 CACTCTGGGTGGAGAGGTCTGCAAGGGAATTGGGGGGGT 1465

RESULT 14
US-09-410-551B-26
Sequence 26, Application US/09410551B
Patent No. 6503737
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.
APPLICANT: REEVES, CHRISTOPHER
APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SAMTI, DANIEL
APPLICANT: WU, KAI
TITLE OF INVENTION: POLYMERIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
TITLE OF INVENTION: CONSTRUCTS THEREFOR
FILE REFERENCE: 30062-20026.00
CURRENT APPLICATION NUMBER: US/09/410,551B
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ. ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 4674
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding syntheitic
OTHER INFORMATION: PKS synthase fragment
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(4673)
US-09-410-551B-26

Alignment Scores:
Pred. No.: 1,01e-06 Length: 4674
Score: 208.00 Matches: 209
Percent Similarity: 31.27% Conservative: 109
Best Local Similarity: 20.55% Mismatches: 400
Query Match: 4.33% Indels: 300
DB: 4 Gaps: 44

US-09-920-705-3 (1-901) x US-09-410-551B-26 (1-4674)
Qy 10 AlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThr 29

Db 1743 GCGAACAAGCGGTGATTCAGAGGGGACCGGAGTGGGTCCCGTGGTATTCGGCCAGG 1802
Qy 30 ThrGluSerAspLeuThrGluLeuPheGlyArgTyr-----Gly 42
Db 1803 ACCGACTCGCTTGTGACTGACGACGAGCGCCGCTTCCGCTGCTATTCGCGCGTCCGCC 1862
Qy 43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
Db 1863 GGGGTGATATGCGGGCTGTGGCATTCGACCTGGGCAATACACGCTGGTGTTCAGACAC 1922
Qy 63 -----ValGluGluAlaValAlaAlaValGluAla 72
Db 1923 CGTCCGCTGCTGTGGAGATGACACCGTCACCGGACCGGACCGCTGTCTGACCTCGGGG 1982
Qy 73 Leu-----GlnGlyAlaAsnLeuAsnGlySerGlnIleValIleGlyTyr 87
Db 1983 GTGTCTGCTTCCCGGACAGGGGTGCGAGCTGTGCTGCGATGGGTGAGAGAACTGCGCGC 2042
Qy 88 AlaArgPro-----AlaAspProCysIleSerLeuTrpValGlyIleGlyProAsn 105
Db 2043 GCGTTCCTCCGCTTCCGCGGATTCATCAGAGTGTG-----GACCTG 2087
Qy 106 ValSerIleAspAspLeuGlu----- 112
Db 2088 CTCGATGTCGCCGATTCGAGGTGAAAGAGACCGGTTACGCCACCGGCGCTTTCGCA 2147
Qy 113 GluGluPheSerIlePheGlyIleValIleGluAspHe----- 124
Db 2148 ATGACAGTGGCTCTGTTCCGCGCTGCTGCAATGCTGGGTGTGTAACACCGACCGGTGATC 2207
Qy 125 ---ArgPheLeuArgGluArgIleGlyThrAlaPheIleAsp---TyrTyrGluMetAspAsp 142
Db 2208 GGCATTCGTGGGTGATGCTTCGCGCTGCTGCAATGCTGGGTGTGTAACACCGGTGATC 2267
Qy 143 Ala-----LeuGlnAlaIleSerMetLeuAsnGlyLeuProMetGlyIleSer 157
Db 2268 GCGTGCATTTGGTGTGTCGCGCGCGGCTCTGATTCAGAGCTTCGCCCGGTTGGGGTG 2327
Qy 158 PheLeuArgValAspPheLeuArgSerGlnAlaProValGlyGluIleTrpAlaGlySer 177
Db 2328 ATGTCGCTGTC----- 2339
Qy 178 TyrAspAsnArgAsnGlyAsnMetLeuHisIleProGlnTyrProHisSerTyrGluAsp 197
Db 2340 -----CCGCTCTCGAGAGATGAG 2357
Qy 198 PheIleGlyAspVal---GlnProSerIleValLeuTrpIleGlyPheProProThrAla 216
Db 2358 GCCCGGCGCTGTGGTGAAGGTGTGAGATCGCGCGGTCAACGCGCGCTGTGCTGCTG 2417
Qy 217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
Db 2418 GTTCTCTCCGATGATGATGACCGCGCTGCGAGCGCGGAGGGGTGGGAAGTGAGAG 2477
Qy 237 ArgValIleSerTyrTrpSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
Db 2478 CGGCTGCGACACGACCGCGCTTCATTCGCGCGGTATGAAA---CCCATGCTGAGAGAG 2534
Qy 257 AlaArgGlnCysLeuGlyLeuGluIleGluGlnIleArgLeuPheAsnAspProGlyIleVal 276
Db 2535 TTCGCGGCGGTGCGCGAAGGCTG-----ACCTACCGACGCGGAGGTCTTCATG 2585
Qy 277 MetTyrSerAsnAspGluLeuProProGlu-----GlnAspAspThr 290
Db 2586 GCGGTGTGATCATGATGACACCGCTGATCTGCTGCTGCGGACGAGTCCGAGACAGCTC 2645
Qy 291 SerPhe-TyrSerGlyMetIleArgSerArgThrAspMetPheAsnAsn----- 306
Db 2646 CCGTTGCGACACAGGTGCGCTGTACGAGACGCGGTGCTGTGAGCTGGGTGCCGAC 2705
Qy 306 ----- 306

Db	2706	CGGTACTGCGCCGCGTGGTTCAGCGGTGTGTGGAGTGTCTGACAGGGGACCCAGCAATTCAG	2765
Qy	307	-----AspProSerCysValSer-----	312
Db	2766	GCCGGATCGCGGCCCTGGCCACCGCTGTATGTCAACGGCTCACGGTGCATGGCGCCG	2825
Qy	313	-----SerProHisSerTh-----	317
Db	2826	CTTCCTGGCGGATGCTCCGGCAACAGGGTGTGAACTTTCGACATACGACTTTCAGAC	2885
Qy	317	rgIyIleProGlySerMet-----ArgProLeuArgGlyThrAsnGluAr	332
Db	2886	CAGCGCTACTGGCTCGAGTCGGGCTCCCGCGCACGGCGCACTCGGGCCACCCCGTCTC	2945
Qy	332	gSerTryAsnGlyAlaGluTryAsnAspValValGly-----	344
Db	2946	GGCACCCGAGTCCCGCTCGCCGGGTGCGCCGGGCGGGGTTCACAGGATCCCGTGGCCGC	3005
Qy	345	-----LysGluProAsnThrArg-----	350
Db	3006	GGTGGCGACCGCGCGGTGTTCATCGCCGAATCGGGCTGCGCGCGCGGACGCCACCGAC	3065
Qy	351	-ArgProSerAlaAsnGlyThrGlyIleleuProSerProThrGlyProGlyIleLeuPr	370
Db	3066	TGGCGCACAGTCTGCAACAGTCTGACCTCATCTCCGTCGCCCGGGATTCGG-----	3114
Qy	370	oSerProAlaGlnGly---ThrArgArgProMetArgSerAsnProAspSerTryGluG	389
Db	3115	-----CCCGCGCAGGGGCCACCGCGCAGACTTGGGTGCATGAACCCGCGCGCGACCGGCGG	3170
Qy	389	Y-----TyrAsnProAlaGlnLeuValArgGluSerIysArgThrArgArgAs	405
Db	3171	CGCCGCTTCACCGTCCACACCGCGTGGCGACCGCCCGGTGACGCTGCACGCGGAGGGG	3230
Qy	405	pgIySerValAspGly-----PheThrProMetGlyValAspG	418
Db	3231	GTTCTCCGCCCGCGCGCGGTGCCCGACGCCAGGCCCTGCACACCGCTGGCCCCCGCGG	3290
Qy	418	uArgSerPhe-----GlyArgGlySerValAlaAlaArgProIleAr	432
Db	3291	GGCGCGGTGCCCGCGACGCGGCTGCCGGGGGGGTGGCGACCGCGGACCAAGTCTTGCTG	3350
Qy	432	gGlyProProAspSerArgPheHisIleThrArgGlyMetIleAlaIysGlyIThrProVa	452
Db	3351	GAAGCCGAAGTCGACA-----GCCCTGACGGCTTCGTGGCACACCCCGCAC	3395
Qy	452	lCysCyAlaArgCysValProMetGlyLysGlyIleGluThrIlysLeuProGluValVa	472
Db	3396	CTGCTCCGACGCGGTCTTCTCCGGG-----TCGGCGAC	3428
Qy	472	IAsnCySerAlaArgThrAspLeuAsnMetLeuAlaIysHisTryAlaValAlaIleG	492
Db	3429	GGGAGCCGCCACGACCGGACCGGATGG- CGGACCTGCGGGTGCAC-----	3470
Qy	492	YCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTryThrGluPh	512
Db	3471	-----GCGTCCGACGCGCACCGT	3487
Qy	512	eLeuArgTryIleuSerSerIysAspArgAlaGlyValAlaIysLeuAsp-----AspG	530
Db	3488	GCTGGCGCCCTGCCCTCACCCCGCGGACAGTGGTGTCTGGAGCTCGCCGCTTTCAGCGG	3547
Qy	530	yThrThrLeu---PheLeuValProProSerAspPheLeuThrAspValLeuGlnValTh	549
Db	3548	TGGCGGATACCGGCTCACCGCGGAGTCCGTGACGCTGGCGGACGCGGTGGCGAGG	3607
Qy	549	rArgGlnGluArgLeuTryGlyValVal-----LeuIysLeuProPro-ProA	565
Db	3608	CGGATCCGACAGTCCGACCGCTGCTCGCTTCGAGTTCAGTTCGCGGTCGCGGAGGACCA	3667
Qy	565	IaValProValThrAlaSerTryIysGlnGluSerGlnSerAsnProLeuHisTryMetA	585
Db	3668	CTACGACGCTCCGACGAGCTGCGCGAGGGCTTACCTCATACCGCGCACACCCCGCA	3727

Oy 565 spglnalAArgAspSerProAlaAsnAlaSerHisSerLeuTyProArgGluAsnT 605
 Db 3728 CGACCCCGACGACCCCAACCCCAACACACACACACACACACACACACACAC 3787
 Oy 605 yrlleArgGlyAlaProGluHisLeuThrAlaAla-----SerIleProSerValSerg 623
 Db 3788 ACGCGCTCCTCAACCGCGCTCTCAACACCACTCATCAACCAACCAACCACTCATGCTCA 3847
 Oy 623 lUpProLeuAlaArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluL 643
 Db 3848 CACCA-----CCACCGACCCCCCAGCGCGCGCTCAACCGGCTCAACCGGACGACCA 3901
 Oy 643 euleAlaAlaThrleuAlaSerIleLeuProAlaThrSergInProAlaAlaProGluSerH 663
 Db 3902 AAAGCAACACCCCGCGCGCA-----TCCACTCATGAAACCAACCAACCCCAACCCCAACCC 3955
 Oy 663 lsgInProMet-----SergIleProSerThr-ValValSerThrAlaHisGlnSerAsn 680
 Db 3956 ACTCCCTCCCTCAACCACTCACACCCCTCCACCAACCCCAACCTTAAGCCCTCAACCAAA 4012
 Oy 681 GlyLeuTyArgAsnGlyAlaAlaProSergInAlaTrpValSArgGlyProGlnThrValHis 700
 Db 4013 -----CACCTCCACACACCCCACTCAACCCCACTCAACCAACCAACCAACCAAC 4066
 Oy 701 AspAlaSerAsnGlnSerPheGlnGlnTyArgGlnGlnTyThrProAlaGlyGlnLeu 720
 Db 4067 AAC-----CACCCCAACACCCCAAC 4087
 Oy 721 ProProProProSerArgTyTyProProAlaSerAsnProAsnTyThrSergIleMet 740
 Db 4088 CCTCAACCCCAACCAACCGCATCTCAT----- 4114
 Oy 741 ValHisGlyAsnMetGlnTyArgGlnSerGlnSerValAsnMetProGlnLeuSerPro--- 759
 Db 4115 ---CACGGGGGGCTTCGGGACCTCGCGGGATCTCGCGGCACTCAACCAACCCCA 4171
 Oy 760 ---LeuProAsnMetProHisAsnAsnTySerMetTyThrGlnGlySerSerAsnHis 778
 Db 4172 CACCTACCTCCTCTCCGCGACACACACACACCCCAACACCGG-----CAC 4219
 Oy 779 ProValSergInProMetValGlnGlnTyArgInProGluAlaSerMetProAsnGlnAsn 798
 Db 4220 CCA-----CATCCCCG- -CGACTCAACGACGACCCCAACCAATCAACCAAGCCCTCAC 4270
 Oy 799 TyrGlyPro-IleProSerTyArgInGlnAlaAsnPheHisGlyValThrThr----- 815
 Db 4271 -----CCACTACCAACACCCCTCACCGGATCTTTCACACCGCGGCACTTCGACGA 4324
 Oy 816 -AsnGlnAlaGlnAsnLeuAsnProSergInPheGlnAlaAlaMetGlnProProAlaAs 835
 Db 4325 CGCCACCCCTCAACCACTCAACCCCAACCTCAACCAACCTTCACCAACCAAGCCGA 4384
 Oy 835 pLyAla-----AsnLeuGluProGlnAsnGlnAlaLeu 846
 Db 4385 CGCGCGCTCGACCTCCACCAACCAACCAACCAACCCCTC 4427
 RESULT 15
 US-09-940-316B-26
 Sequence 26, Application US/09940316B
 Patent No. 6759536
 GENERAL INFORMATION:
 APPLICANT: KOSAN BIOSCIENCES, Inc.
 APPLICANT: REEVES, CHRISTOPHER
 APPLICANT: CHU, DANIEL
 APPLICANT: KHOSLA, CHAITAN
 APPLICANT: SANTY, DANIEL
 APPLICANT: WU, KAI
 TITLE OF INVENTION: POLYKETIDES ENCODING THE FKBA GENE OF THE FK-520 POLYKETIDE SYNTHASE
 FILE REFERENCE: 30062-20026.11
 CURRENT APPLICATION NUMBER: US/09/940.316B
 CURRENT FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: 09/410,551
PRIOR FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ. ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ. ID NO 26
LENGTH: 4674
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
OTHER INFORMATION: PKS synthase fragment
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(4673)
US-09-940-316B-26

Alignment Scores:
Pred. No.: 1,01e-06 Length: 4674
Score: 208.00 Matches: 209
Percent Similarity: 31.27% Conservative: 109
Best Local Similarity: 20.55% Mismatches: 400
Query Match: 4.33% Indels: 300
DB: 4 Gaps: 44
US-09-920-705-3 (1-901) x US-09-940-316B-26 (1-4674)
QY 10 AlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThr 29
DB 1743 GCGCAACACCGCGTATCGAGCGGCGCACCGAGTGGTGGCGTGGATTCGCCAGG 1802
QY 30 ThrGluSerAspLeuThrGluLeuPheGlyArgTyr-----Gly 42
DB 1803 ACCGAGTCGGCTTGTACTGACGACGAGGCGCGTGGCGTATCTGGCGCGTGGCC 1862
QY 43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
DB 1863 GGGGTGATATGGGGGTGTGGCATCGACCTGGCGCATGACAGGTGGGTTCGAGCAC 1922
QY 63 -----ValGluGluAlaValAlaAlaValGluAla 72
DB 1923 CGTGGCGTGTGGGAGATGACACCGTACCGGCGCGCTGTGTGTGACCTCGGCGC 1982
QY 73 Leu-----GlnGlyAlaAsnLeuAsnGlySerGlnIleValIleGluTyr 87
DB 1983 GTTTCGTCTTCCCGGACAGGGGTGCGACCGTGTGTGATGGTGAAGACTGGCGCC 2042
QY 88 AlaArgPro-----AlaTyrProCysLysSerLeuTrpValGlyIleGlyProAsn 105
DB 2043 GCGTTCCTCCCTCTTCCGGCGGATCCATCAGCAGGTGCG-----GACCTG 2087
QY 106 ValSerLysAspAspLeuGlu-----GACCTG 2087
DB 2088 CTCGATGTGCGCATCTGAGAGTGAAGACAGCGGTTACCGCCAGCGCGCTGTTCGA 2147
QY 113 GluGluPheSerLysPheGlyLysIleGluAspPhe-----124
DB 2148 ATGACAGTGGCTCTGTTCGGGCTGTGTAATCGTGGGTGTACGACCGAGCGGTGATC 2207
QY 125 ---ArgPheLeuArgGluArgLysThrAlaPheIleAsp---TyrTyrGluMetAspAsp 142
DB 2208 GCGCATTCGGTGGTGAAGCTTGGCGCTGCGCTATATGTCCGGGTGTGTGTGAGAGAT 2267
QY 143 Ala-----LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySer 157
DB 2268 GCGTCGACATTGTGTGTGGCGGCGGTCTGTGTATGACAGGCTGCGCGGAGGAGGT 2327
QY 158 PheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySer 177

DB 2328 ATGCTGCTGTC-----2339
QY 178 TyrAspAsnArgAsnGlnLysMetAsnHisLysProGlnTyrProHisSerTyrGluAsp 197
DB 2340 -----CCGCTCTCGAGAGATGAG 2357
QY 198 PheLysGlyAspVal---GlnProSerLysValLeuTrpIleGlyPheProThrAla 216
DB 2358 GCGCGGCGCGTGTGGGTAGAGGTGTGAGATGCGCGGTCAACGCGCGTGTGCGTGC 2417
QY 217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyLysIleGlu 236
DB 2418 GTTCTCTCCGGTATAGAGCGCCCGCTGTGACGCGCGGAGGCGTGGGAGATGAGAC 2477
QY 237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
DB 2478 GCGCTGGCAGCAGCAGCGCGTTCATTCGCGCGTATGGA---CCATGTCTGAGAGAG 2534
QY 257 AlaArgGlnCysLysGluGluLysLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 276
DB 2535 TTCCGGCGCGTGTGCGCAGAGCGCTG-----ACCTACCGAGCGCGAGTCTCATG 2585
QY 277 MetTyrSerAsnAspGluLeuProProGlu-----GlnAspAspThr 290
DB 2586 GCGCTGTGTATAGGTGACACCGCTGTGATGCTGTGCGGAGTCCGGACAGCGTC 2645
QY 291 SerPhe-TyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn-----306
DB 2646 CGGTCGCGAGAGAGGTGCGCTGTACAGAGACCGCTGTGTGAGTGGGTGCGGAC 2705
QY 306 -----306
DB 2706 CGGTCACTGACCGCGCTGTGACAGGTGTGCGGATGCTGACGCGACGACGAAATCAG 2765
QY 307 ---AspProSerCysValSer-----312
DB 2766 GCGCGATGCGCGCGCTGTGACCTGTATGTACAGCGGTCAAGTCACTGACGCGCGC 2825
QY 313 -----SerProHisSerThr 317
DB 2826 CTCCTGGCGAGTCTCGGCAACAGCGGTGTGACCTTCCGACATACCTTCCAGCAC 2885
QY 317 GlnIleProGlySerMet-----ArgProLeuArgGlyThrAsnGluArg 332
DB 2886 CAGCGTACTGTGCTGATGTGCTCCCGCGCACCGCGCTGCGCGCACCGCGCTCTC 2945
QY 332 GSerTyrAsnGlyAlaGluTyrAsnAspValValGly-----344
DB 2946 GGCACCGAGTGCCTGCGCGGCTGCGCGGCGGCGGTTCACGAGTCCCGTCCCGCC 3005
QY 345 -----LysGluProAsnTrpArg-----350
DB 3006 GGTGGGAGACCGCGGTGTTCATCGCGAATCGGCGCTGCGCGCGCGCGACCGCACGAC 3065
QY 351 -ArgProSerAlaAsnGlyThrGlyLysLeuProSerProGlnGlyProGlyLysLeuP 370
DB 3066 TGGCCACGATCGAAGAGCTGACGTCACTGCTGCTGCGCGCGAGTCCG-----3114
QY 370 oSerProAlaGlnGly---ThrArgArgProMetArgSerAsnProAspSerTrpGluG 389
DB 3115 ---CCCGGCGAGGCGCACCGCGACCTGTGCTGATGAAACCCCGCGCGAGCGGCG 3170
QY 389 Y-----TyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAs 405
DB 3171 CGCGGCTTACCGTTCACACCGCGGTGCGGAGCGCGCGTGAACGTGACCGCGAGGG 3230
QY 405 pGlySerValAspGly-----PheThrProMetGlyValAspG 418
DB 3231 GTTCTCGCGCGCGCGCGCTGCGCGACCGCGTGCACACCGCTGCGCGCGCGCGCG 3290
QY 418 wArgSerPhe-----GlyArgGlySerValAlaAlaArgProIleArg 432

```

Db      3291 GGGCGGCTGCCCGGAGCGGGCTGCCCGGCGGAGCGCGGAGCAGGCTTCCTGTC 3350
QY      432 GGGYProProAspSerAspHisIleTrrArgGlyMetIleAlaGlyGlyThrProVa 452
Db      3351 GAAGCCGAAGTCGACA-----GCCCTGACGGCTTCGTCGACACCCCGAC 3395
QY      452 IcySyaAlaArgCyValPromeGlyGlyIleGluThrTyLeuProGluValVa 472
Db      3396 CTGCTGACCGCGGCTTCCTCCGG-----TCGGCGAC 3428
QY      472 IAsnCySerAlaArgThrAspLeuAsnMetLeuAlaGlyHisIleTrrAlaIleG 492
Db      3429 GGGAGCGCGCAGCGGACCGGATG3-CGCGACCTCGCGTGCAC----- 3470
QY      492 yCySgIuIleValPhePheValProAspArgGluGluAspPheAlaSerTyThrGluPh 512
Db      3471 -----GCCCTCGGACGCCACCGT 3487
QY      512 eLeuArgTyrLeuSerSerIysAspArgAlaGlyValAlaIleLeuAsp-----AspG 530
Db      3488 GCTGGCGGCTGCTGCTACCGCGCGGACAGTGGTCTGAGAGCTGCCGCTTCGACG 3547
QY      530 yThrThrLeu--PheLeuValProProSerAspPheLeuThrAspValIleuGluValTh 549
Db      3548 TGCCGGAATGCCGCTGCTCACCGCGAGTGTGAGTGCAGTGGCGAGTGCCTGCGCAG 3607
QY      549 rArgGluArgLeuTyrglyValVal-----LeuTyLeuProPro--ProA 565
Db      3608 CGGATCCGAGAGTCGAGCGCTGCTGCTGCGCTTGAGTGGTGGCGGAGGCCA 3667
QY      565 IValProValThrAlaSerTyArgGlnIuSerIuSerIuAsnProLeuHisIleTrrMetA 585
Db      3668 CTACGACGGGTGCGAGAGTGCAGGGCTACACCTCTATACCGCCACACACCCGA 3727
QY      585 spGlnAlaArgAspSerProAlaAsnAlaSerHisSerIeutyTyProArgGluAsnT 605
Db      3728 CGACCCGAGAGCCCAACCCCAACACCAACCCACACCGCACCAACAAACAC 3787
QY      605 yTrIleArgGlyAlaProGluHisLeuThrAlaAla-----SerIysProSerValSerG 623
Db      3788 ACCGCTCTACCGGCTTCACACACACCTCATCACACCAACACACCTCATGCTCA 3847
QY      623 IuProLeuArgIleProAsnAsnAlaIaProGlnAlaGlyValSerIeuthrProGluL 643
Db      3848 CACCA-----CCACCGACCCCGGCGCGGCTCACCGGCTCACCGCGACCGCACA 3901
QY      643 euleuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerH 663
Db      3902 AAACGAACCCCGCGCA-----TCACCTCATCGAAACCCACACCCCGACCC 3955
QY      663 IeGlnProMet-----SerGlyProSerThr--ValIleSerThrAlaHisGlnSerAsn 680
Db      3956 ACTCCCCCTACCCCACTACCACTCTCACCAACCCCACTACGCTTCACACACA 4012
QY      681 GlyLeuTyArgnIyGluAlaProSerGlnAlaTrrIyArgGlyProGlnThrValHis 700
Db      4013 -----CACCTTCACACCCCGACCTCACCCCATCACACCAACACCAACAC 4066
QY      701 AspAlaSerAsnGlnSerPheGlnGlnTyrglyAsnGlnTyThrProAlaGlyGlnLeu 720
Db      4067 AAC-----CACCCCGACACCCCGAC 4087
QY      721 ProProProProSerArgTyrrProProAlaSerAsnAsnProAsnTyThrSerGlyMet 740
Db      4088 CTTCAACCCCAACCAACGCACTCTCAT----- 4114
QY      741 ValHisGlyAsnMetGlnTyrglnSerGlnSerValAsnMetProGlnLeuSerPro-- 759
Db      4115 ---CACCGGCGGCTCGGACCTCGCGCATCTCGCGCGCACCTCAACACCCCGCA 4171
QY      760 ---LeuProAsnMetProHisAsnAsnTyrrSerMetTyThrGlnGlySerSerAsnHis 778
Db      4172 CACCTACCTCTCTCTCCGACACCAACCCCGACACCCCGACACACCCCGG-----CAC 4219

```

```

QY      779 ProValSerGlnProMetValGlnGlnTyrglnProGluAlaSerMetProAsnGlnAsn 798
Db      4220 CCA-----CATCCCTG---CGACCTCACGACCCCAACCCCAAAATCACCCAGCCCTCAC 4270
QY      799 TyrglyPro-IleProSerTyrglnGlnAlaAsnPheHisGlyValThrThr----- 815
Db      4271 -----CCATATACCAACCAACCCCTCACCGGATCTTCCACACCGCGCCGCTCGACGA 4324
QY      816 -AsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaIleMetGlnProProAlaAs 835
Db      4325 CGCACCTTCACCAACCTTCACCCCGCAACCTCACACCTCACACCAACCTTCACCAACCAAGCCGA 4384
QY      835 pIySAla-----AsnLeuGluProGlnAsnGlnAlaLeu 846
Db      4385 CGCGGCTTGGACCTTCACACCAACCCCGACCAACCCCGCTC 4427

```

Search completed: April 8, 2005, 13:52:02
 Job time : 453 sec8

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 12:01:35 ; Search time 1224 Seconds

(without alignments)
4460.836 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801
Sequence: 1 MALSMKPRRDSGFSNNL.....LQIQKQKQSSGSTRACGP 901

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5615251 seqs, 3030001701 residues

Total number of hits satisfying chosen parameters: 11230502

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=x1h
-Q=/cg2_1/USPTO.epool.h/US09920705/runat_07042005_101816_20810/app_query.fasta_1.1095
-DB=Published Applications NA -QFMT=fastp -SUFFIX=rmh -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdt -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09920705 @CGN_1_1_622 @runat_07042005_101816_20810
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCL=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: *
1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
21: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
22: /cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4801	100.0	2706	10	US-09-920-705-2
2	4392	91.5	4593	10	US-09-920-705-1
3	1436	29.9	801	10	US-09-920-705-4
4	1286.5	26.8	3453	18	US-10-425-115-123950
5	1281.5	26.7	3084	18	US-10-437-963-36321
6	802	16.7	3315	17	US-10-437-963-79215
7	669	13.9	1101	18	US-10-424-599-59704
8	651.5	13.6	3715	10	US-09-920-705-6
9	499	10.4	801	17	US-10-424-599-65572
10	466	9.7	349	10	US-09-920-705-5
11	463	9.6	1630	17	US-10-425-114-28888
12	447	9.3	1631	17	US-10-424-599-40095
13	406	8.5	2159	17	US-10-425-114-31541
14	382	8.0	815	18	US-10-767-701-2544
15	359	7.5	606	18	US-10-767-701-30057
16	336.5	7.0	6836	18	US-10-475-917-1
17	333.5	6.9	1146	17	US-10-424-599-18149
18	309	6.4	729	17	US-10-424-599-47401
19	302	6.3	483728	18	US-10-699-156-2
20	297	6.2	1621	18	US-10-425-115-118399
21	276.5	5.8	3383	18	US-10-475-917-11
22	273	5.7	3312	18	US-10-475-917-7
23	271.5	5.7	10656	17	US-10-263-929-42
24	270	5.6	1241	17	US-10-425-114-35702
25	268	5.6	3228	17	US-10-302-172-752
26	265	5.5	10995	17	US-10-263-929-41
27	265	5.5	12227	15	US-10-177-293-422
28	265	5.5	12227	18	US-10-684-422-243
29	261	5.4	11295	10	US-09-814-353-15153
30	258.5	5.4	683	17	US-10-424-599-2550
31	254	5.3	438	9	US-09-770-444-972
32	246.5	5.1	1871	9	US-09-834-975-1036
33	246	5.1	2228	17	US-10-172-118-410
34	246	5.1	2228	17	US-10-342-887-410
35	246	5.1	2482	18	US-10-723-860-6794
36	241.5	5.0	443	18	US-10-425-115-79600
37	236.5	4.9	8147	13	US-10-109-886-9
38	229.5	4.8	1803	9	US-09-938-842A-1801
39	229.5	4.8	1803	11	US-10-263-929-40
40	225	4.7	16683	17	US-10-425-115-118394
41	223.5	4.7	2201	18	US-10-425-115-31949
42	218.5	4.6	2340	18	US-10-723-860-8366
43	216	4.5	2774	18	US-10-425-115-25600
44	212	4.4	3433	18	US-10-425-115-25600
45	212	4.4	3433	18	US-10-425-115-25600

ALIGNMENTS

RESULT 1
US-09-920-705-2
Sequence 2, Application US/09920705
Publication No. US20030079252A1
GENERAL INFORMATION:
APPLICANT: Amasino, Richard M.
APPLICANT: Schomburg, Fritz M.
APPLICANT: Michaels, Scott D.
APPLICANT: Patton, David
TITLE OF INVENTION: Floral Induction Gene
FILE REFERENCE: 960296. 97214
CURRENT APPLICATION NUMBER: US/09/920, 705
CURRENT FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2706
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:

NAME/KEY: CDS
LOCATION: (1) .. (2706)
NAME/KEY: misc binding
LOCATION: (46) .. (279)
OTHER INFORMATION: RNA Binding Region
NAME/KEY: misc binding
LOCATION: (283) .. (522)
OTHER INFORMATION: RNA Binding Region
NAME/KEY: misc binding
LOCATION: (610) .. (852)
OTHER INFORMATION: RNA Binding Region
US-09-920-705-2

Alignment Scores:

Pred. No.:	0	Length:	2706
Score:	4801.00	Matches:	901
Percent Similarity:	100.008	Conservative:	0
Best Local Similarity:	100.008	Mismatches:	0
Query Match:	100.008	Indels:	0
DB:	10	Gaps:	0

US-09-920-705-3 (1-901) x US-09-920-705-2 (1-2706)

QY 1 MetAlaLeuSerMetLysPProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
DB 1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATCCGGTTTCCAGTCAAAATCTT 60
QY 21 TrpValGlySerLeuThrProGlnThrGlnSerAspLeuThrGlnLeuPheGlyArg 40
DB 61 TGGGTGGTGGTCCCTAAAGCCGAGACGACGAGTCAAGTCAATCTGACCGAGTGGTTGGAGA 120
QY 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
DB 121 TAGGGGAGATATGATGAATCAAGGTGATCTTCTTCAAGAGCTTCCGTTATATATCTAC 180
QY 61 ArgHisValGlnGlnAlaValAlaAlaLysGlnAlaLeuGlnGlyValAsnLeuGngly 80
DB 181 AGACATGTGGAGAGACAGTCGACGACCAAGAGGCTCTTCAAGACAAATTTGAATGGA 240
QY 81 SerGlnIleLysIleGlnTyrAlaArgProAlaLysProGlyLysSerLeuThrValGly 100
DB 241 AGCAAAATTAAGTTCATTAACGACGACGACGACGACGACGACGACGACGACGACGACG 300
QY 101 GlyIleGlyProAsnValSerLysAspAspLeuGlnGlnLysPheSerLysPheGlyLys 120
DB 301 GGAATCGCGCTTAATGCTCTCAAGATGACTGAGAGAGAGATTCAGCAAGTTTGGGAAA 360
QY 121 IleGlnAspPheArgPheLeuArgGlnLysGlnThrAlaPheIleAspTyrTyrGlnMet 140
DB 361 ATCGAGGATTTTAAAGTTCTTCAGAGAACCAAGACGCTTTCATTGATTAATATGAGATG 420
QY 141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
DB 421 GATGATGCTTTACAGCTTAAGCATGAATGAAAGCTTAAGGATGATGCTTTTGGCT 480
QY 161 ValAspPheLeuArgSerGlnAlaProLysLysGlnGlnThrAlaGlySerTyrAspAsn 180
DB 481 GTTGAATTTCTCCGGTCAAGGCGCAAAAGAAACAATGGCGTGGCTCTTAACGATAC 540
QY 181 ArgAsnGlyLysMetAsnHisLysProGlnTyrProHisSerTyrGlnAspPheLysGly 200
DB 541 AGAAATGGCAATATGATATAAACCGCAGTATCTCACTCATATGAAACCTTTAAAGGA 600
QY 201 AspValGlnProSerLysValLeuThrIleGlyPheProProThrAlaThrGlnCysAsn 220
DB 601 GATGTCAGCAAGTAGAGTTCTTGAGATGGGATTCCTCTTCAAGCTACACAAATGCAT 660
QY 221 AspGlnGlnIleLeuHisAsnAlaMetIleLeuPheGlyGlnIleGlnArgValLysSer 240
DB 661 GATGAGCAAAATCTGCAACATGAGATGATCTTTGGTGAATGAGAGGGTAAAGT 720
QY 241 TyrProSerArgAsnPheAlaLeuValGlnPheArgSerAlaGlnGlnLys 260

DB 721 TACCCATCAAGGAATTTTGCACCTTGAGATTTAGAGCGCGAGAACTCGCAATGC 780
QY 261 LysGlnGlyLeuGlnGlyArgLeuPheAsnProArgIleLysIleMetTyrSerAsn 280
DB 781 AAGGAAGGCTCAAGGGGAGTTATCAATAATCTTAATCAAAATATATGACTCAAC 840
QY 281 AspGlnLeuProGlnGlnAlaAspAspPheTyrSerGlyMetLysArgSerArg 300
DB 841 GATGATGCTCTCTGAGCAGACGATCTAGTTTTCCTGCTATGATAAAGCGTCAAG 900
QY 301 ThrAspMetPheAsnAspAspProSerCysValSerSerProHisSerThrGlyIlePro 320
DB 901 ACAGATATGTTCAATATATGATCTTCAATGATTTTCTCTCATCTACATCGAATTCCT 960
QY 321 GlySerMetArgProLeuArgGlyThrAsnGlnLysSerTyrAsnGlyValGlnTyrAsn 340
DB 961 GGTGTATAGAGCCCTCCAGAGTACGATGAGTCAAGCTTATATATGTCAGAAATACAT 1020
QY 341 AspValValGlyLysGlnProAsnThrPheArgProProSerAlaAsnGlyThrGlyIleLeu 360
DB 1021 GACGTTGTTGGTAAAGACCAACTGAGAGAGGCCATTCGCAATGGAATGGAATCTC 1080
QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
DB 1081 CCATCTCCAAAGAGACTGGAATCTCCATCTCTGCAACAGGTAGAGGCGGCTTATG 1140
QY 381 ArgSerAsnProAspSerTyrGlnGlyTyrAspProAlaGlnLeuValArgGlnSerLys 400
DB 1141 AGGTCAAAACCCGATCTTGGGAAGATATATCTCTGATGTTGTCAGAGAAAGTAAA 1200
QY 401 ArgThrArgAspArgGlySerValAspGlyPheThrProMetGlyValAspGlnLysSer 420
DB 1201 CGAACCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
QY 421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
DB 1261 TTTGGTCAGAGTTCAAGTTGCTGATGACCTATCCGTGGGCCCTGATTCGATCACATA 1320
QY 441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMet 460
DB 1321 TGGAGAGGAATGATTTGCGAAGGTGGAATCTCCGCTGTTGCTGCTGTTGCTGCTG 1380
QY 461 GlyLysGlyIleGlnThrLysLeuProGlnValValAsnCysSerAlaArgThrAspLeu 480
DB 1381 GGAAGAGGATTTGAATCAATCACTGAGTCTCAATGTTTTCAGCAAGACATGATTTG 1440
QY 481 AsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlnIleValPhePheValPro 500
DB 1441 AATATGCTCGCTAAACATTAAGCGGTGCAATGATGATGATGATGATGATGATGATG 1500
QY 501 AspArgGlnLysAspPheAlaSerTyrThrGlnPheLeuArgTyrLeuSerSerLysAsp 520
DB 1501 GACAGGAGGAAGATTTTGGCTTTACCTGAATTTCTCCGATACCTTAAGCTCAAAAGAT 1560
QY 521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
DB 1561 CGGGCGGGTGTGGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 541 PheLeuThrAspValLeuGlnValThrArgGlnGlnLysGlnLysValValLeuLys 560
DB 1621 TTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
QY 561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGlnLysGlnLysAspPro 580
DB 1681 TTACCCCGGACAGCGCTTCTGTTACAGATCAATCAAGACAAAGATTCAGTCCAAATCCT 1740
QY 581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
DB 1741 CTGATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
QY 601 ProArgGlnLysTyrIleArgGlyValaProGlnHisLeuThrAlaLysSerLysProSer 620
DB 1801 CTTAGGAGAAATTAATTAAGGGGTGACAGAAATTTGACAGCTGCTCAAAACATCT 1860

QY 621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
DB 1861 GTTAGGAGGCTCTCAAGATTAATTAATGACAGCGCTCAAGCTGGGGTTAGTTACT 1920
QY 641 ProGlnLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
DB 1921 CCGAGGCTTTTAGCACTCTGGCATCTATCTCCCTGCAACTCTCAACTGCTGCCCT 1980
QY 661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
DB 1981 GAGAGTCAACCACTATGTGACAGACCTTCAACAGTTGTTCCACAGACATCATGCTCAAT 2040
QY 681 GlyLeuTyraAsnGlyGlnAlaProSerGlnAlaAlaTrpLysArgGlyProGlnThrValHis 700
DB 2041 GGACTGTACATGAGAGACACGCTCAAGCTTGAAAAGAGTCCAAACAGATTCTAT 2100
QY 701 AspAlaSerAsnGlnSerPheGlnGlnTyrglyAsnGlnTyrrProAlaGlyGlnLeu 720
DB 2101 GATGCGTCAATCATCATTCCTCAACAAATACGAAATACATCTCCAGCTGGGCAACTA 2160
QY 721 ProProProProSerArgTyrrProProAlaSerAsnAsnProAsnTyrrThrSerGlyMet 740
DB 2161 CCTCTCTCTCTCGCGTTACCTCCAGCTTCAAAACCCCAACTACATGATGAGATG 2220
QY 741 ValHisGlyAsnMetGlnTyrglyAsnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
DB 2221 GTTCATAGGCAACATGCAATACGAGAGCAATCTGTTAACTGCTCAAGCTGCTCGCTA 2280
QY 761 ProAsnMetProHisAsnAsnTyrrSerMetTyrrThrGlnGlySerSerAsnHisProAla 780
DB 2281 CCAAAATATGCTCTATATATATTTTCCATGATACATCAGGGTTCGTCAATCATCTCTGT 2340
QY 781 SerGlnProMetValGlnGlnTyrglnProGlnAlaSerMetProAsnGlnAsnTyrgly 800
DB 2341 TCTCAGCCCATGCTCCGCAATACCAACGAGCGCTTCCAAACCAAACTATGCT 2400
QY 801 ProAlaProSerTyrrGlnGlnAlaAsnPheHisGlyValThrTrpAsnGlnAlaGlnAsn 820
DB 2401 CCAATTCGAATGATACGAGGCTAATTTTTCATGCGCTAACCAATCGGACAGAAC 2460
QY 821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuAla 840
DB 2461 TTAAACCTTCCCAATTTCAAGCTGCCATGCCAACACACAGATTAAGGAAATTTTAGG 2520
QY 841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
DB 2521 CCACAAAACCAAGCACTACGATTCGAGCTATGATCTCTGGGATGGTCAAGGTTACACA 2580
QY 861 AspGlyGlyValAspLysAsnGlnArgTyrrGlnSerThrLeuGlnPheAlaAsnLeu 880
DB 2581 GATGGGAGGTGATGATGAATGAGATGACAGATACAGCTCAACATTTTGGACAAACCTT 2640
QY 881 LeuLeuGlnIleGlnGlnTyrglnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
DB 2641 CTTCTCCAGTATACGAGAAACGACAGCAACATCTTCAAGTACTCGGCTGACAGGGG 2700
QY 901 Pro 901
DB 2701 CCT 2703

RESULT 2

US-09-920-705-1
; Sequence 1, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaels, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296, 97214
; CURRENT APPLICATION NUMBER: US/09/920, 705

US-09-920-705-3 (1-901) x US-09-920-705-1 (1-4593)
CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4593
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-920-705-1
Alignment Scores:
Pred. No.: 0 Length: 4593
Score: 4392.00 Matches: 899
Percent Similarity: 58.72% Conservative: 0
Best Local Similarity: 58.72% Mismatches: 2
Query Match: 91.48% Indels: 631
DB: 10 Gaps: 4
QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
DB 1 ATGGCGTTATCTATGAAAGCCATTCAAGCGATGATTCGGTTTCCAGTCAACAACTCTT 60
QY 21 TPValGlySerLeuThrProGluThrGlnGlnSerAspLeuThrGluLeuPheGlyArg 40
DB 61 TGGGTGGTAACTTAAAGCCGCGAGACGACAGAGTCAATCTACCGAGTTGTTGAGAA 120
QY 41 TyrglyAspIleAspArgIleThrValTyrrSerSerArgGlyPheAlaPheIleTyrr 60
DB 121 TACGGGATATTGATGATGATCAAGGATGATCTTCAAGAGGCTTTCGTTTATATCTAC 180
QY 61 ArgHisValGlyGlnAlaValAlaAlaLysGlnAlaLeuGlnGlyValAsnLeuAsnGly 80
DB 181 AGACATGTGAGAGAGACAGTCCGACCAAGAGGCTTTCAAGAGCAAAATTTGAATGA 240
QY 81 SerGlnIleLysIleGlyTyrrAlaArgPro----- 90
DB 241 AGTCAAATTAATGCAATACGACAGACCGGTTTGTCTTATCTATCTCTGTTGTTTC 300
QY 90 ----- 90
DB 301 TCTAATTTGATGTTCTTTTGTCAAGATTATCTTTTTCGAATCATAGTCCAGGT 360
QY 90 ----- 90
DB 361 TCACAAACTTGAATGATGCTTTGTTAGTCAAAAATTTCTTGTAATCTGTTTTTTT 420
QY 90 ----- 90
DB 421 TTCCTCATGTACAAATCAAACTGCAAACTGTTTTTTTCTATATATACGTGTTAGCTT 480
QY 90 ----- 90
DB 481 AAGGCAAACTGATCCGATCGAAAGCTTTTCTCAATTACTTTGGTTATATGAACT 540
QY 90 ----- 90
DB 541 CCGGCAAAACCAACCAACAGAGAGCTTCAAAAATTTGATGTTAAAGCATATATACTC 600
QY 90 ----- 90
DB 601 TTAGCAATAGCTCTGCAAAAAGAAATACATCAACACATGTTTACTCTCGTTTATGCGAA 660
QY 90 ----- 90
DB 661 GAAGGTTTAATCCGATTTGCTTTATCTGCAAACTTGTATCGCGTTTACGATGTGATT 720
QY 90 ----- 90
DB 721 TCAGGTGCTTCAGATGTTAATCTCAACAGCTTGAATGAATGATTTGGTACTCA 780
QY 90 ----- 90

Db 781 CTGAAATTGCAATGCTGAAAAATTGGAAGCAGCAAACTATCATCGGCCAAACAGA 840
 QY 90 ----- 90
 Db 841 TAAGAAATTTGGAGTTAAAGTTCAAGTTCTCGCAAAAATCAACCCTGGAGAAAT 900
 QY 90 ----- 90
 Db 901 TTGTCTATGGCGGTAGCTAGATATCAATACGTGATCTGGAAAACAGAAAATTTCTGCA 960
 QY 90 ----- 90
 Db 961 CCGTATGCTCTAATCTTATGATTCGCTGACATATCCATATCAGTGTGAGGAGTTT 1020
 QY 90 ----- 90
 Db 1021 GGGCTTGTGGCTGATGCTTCAAGAAACAATTAAGTGTGTAAAGCGGCTAAC 1080
 QY 90 ----- 90
 Db 1081 TCTACCAATGAAAGCCTTATTCGAAGAACATGTTGTCTCAATTCACATCCCTA 1140
 QY 90 ----- 90
 Db 1141 CGTACAATCTGGGCTTCCCATTTAGTTCCTTAGGAAGTTGACTTCTTCAACACCATTC 1200
 QY 90 ----- 90
 Db 1201 TGGGATTTGGACAGTTTGCAGTGAACAACAAATTCATTTGACAGCAGACTCAATTT 1260
 QY 90 ----- 90
 Db 1261 CAATGCTTGGCGGCTTGTAGTACTTAAATGAGTGTTCAGGATTTGTTATTGGGAC 1320
 QY 90 ----- 90
 Db 1321 TCAGCAATGATTAAGAACAGAAATGTTTGCAAGAAATTAATGCTTTCCGCGCTTG 1380
 QY 90 ----- 90
 Db 1381 GGGACTACATCTTCTCTACAGCAAAATCTGATGTTTAGGACAGAAACTATTACA 1440
 QY 90 ----- 90
 Db 1441 CTGATTAAATAGAAAAAGAGAGATGTTCTCTCGCAATTTTATTGTTAGAA 1500
 QY 90 ----- 90
 Db 1501 AGCAATTGATGAATGCTGTCTGATGCTAGTTGAATTAAGTTAGTTGTGTGTT 1560
 QY 90 ----- 90
 Db 1561 AGTTCTTGATGTTGATGCTTTATCTTGCAACCTATCTGGGATGATGCCCTTT 1620
 QY 91 ----- AAlaLysProCysLysSerLeuTyrValGlyIleGly 103
 Db 1621 ATCGACCTTTCTTGTGGCAGGCAAAACCTGTAAAGCTATGGGTGGGTGAATCGGC 1680
 QY 104 ProAsnValSerLysAspAspLeuGluGluPheSerLysPheGlyValIleGluAsp 123
 Db 1681 CCAATGCTCCCAAGATGACCTGGAGAAAGTTCAAGAAATTTGGGAAAAATCGAGAT 1740
 QY 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
 Db 1741 TTATAGTTCTCGAAGAACGCAAGACGCTTCAATTATTAAGATGATGATGCT 1800
 QY 144 LeuGlnAlaLysSerMetLysGlyLysPheMetGlyLysPheLeuArgValAspPhe 163
 Db 1801 TTACAGGCTTAAGACATGATGAAGAAAGCTATGGGTGATGCTTTTGGGTGATTTT 1860
 QY 164 LeuArgSerGlnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 171
 Db 1861 CTCGGTCAAGAGCCCAAAAAGTAAGCACTTTGTGGCAATTGATTTTACTTTTGA 1920

QY 171 ----- 171
 Db 1921 AAAGCTCCAGTAAACATTTGTTAGTTCAATAATTTGGCTCAAACTGATAGGCTGAG 1980
 QY 171 ----- 171
 Db 1981 CTCTGTTGTGCCCTAGAGCACTATTACTGCTCTATTTCATTTGATGAGTGGCT 2040
 QY 172 ----- GluGlnTyrAlaGly 176
 Db 2041 CAATCTTAAAGTCAAAATCAAGTTACCTTTGTGTATCTTCAGAAACAATGGGCTGCG 2100
 QY 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysPheProGln----- 190
 Db 2101 TCTTACGATACAGAAATGGCAATATGATCAATAAACCGCA-GGTTAGCTTGAATGTTG 2159
 QY 190 ----- 190
 Db 2160 AAAGTATGCTCTGTCTAGTGAATATGATATAGTTACTGTGTTGACGTTTGTATTA 2219
 QY 191 ----- TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
 Db 2220 TTCTTACAGTATCTCTCACTCATATGAAAGCTTTAAAGAGATGTCCAGCCAAAGTAAAGGCT 2279
 QY 208 LeuTyrPileGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
 Db 2280 CTGTGATTTGGGTTCTCTCTCTACTGCTACACAAATGCAATGATAGCAAAATTTGCAACAT 2339
 QY 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
 Db 2340 GCGATGATATCTTGTGATGATGAGAGGGTAAAGATTAACCATCAAGAAATTTTGCA 2399
 QY 248 LeuValGluPheArgSerAlaGluLysValArgGlnCysLysGlyLysGluGlnIleValArg 267
 Db 2400 CTGTGGAGTTTAGAGCGCGAGAGAGCTGCCAATGCCAAGAGGCTTACAGGGGAGG 2459
 QY 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProGluGln 287
 Db 2460 TTATTCAATATCTCGAATCAAAATTAATGATCTCAACAGATGATGCTCTCTGAGCA 2519
 QY 288 AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsp 307
 Db 2520 GACGATACTAGTTTACTCTGATGATAAAGGTCAGAGACAGATATGTTCAATTAATGAT 2579
 QY 308 ProSerCysValSerSerProHisSerThrGlyLysProGlySerMetArgProLeuArg 327
 Db 2580 CTTCAATTGTATCTTCTCTCATTTCACTGAATTCCTGGGTATAGAGGCCCTCAGA 2639
 QY 328 GlyThrAsnGluArgSerTyrAsnGlyValIleGluTyrAsnAspValIleGlyLysGluPro 347
 Db 2640 GGTACAAATGAGCGCTCATTAATATGTCGAGAAATACATGACGTTGTTGGTAAGAGCA 2699
 QY 348 AsnTyrPargArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
 Db 2700 AACTGAGAGAGCGCATCTGCAATGAACTGAATATCTCCATCTCCAAACAGACCTGGA 2759
 QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyr 387
 Db 2760 ATCCCTCCATCTCTGCAAGAGTACAGAGGCGCTTAAGAGTAAACCCCGATCTTGG 2819
 QY 388 GluGlyTyrAspProAlaGlnLeuValArgLysSerLysArgThrArgArgAspGlySer 407
 Db 2820 GAAGATATGATCTCTGCTCACTGTGTGTGTCAGAGAAATGAACACACAAAGAGATGATCA 2879
 QY 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
 Db 2880 GTGACGCTTTTACTCCAAATGGGTGTGATGACAGGTCATTTGTGTCAGGTTCAAGTGTCT 2939
 QY 428 AlaArgProIleArgGlyProProAspSerAspHisIleTyrPargGlyMetIleAlaLys 447
 Db 2940 GCTAGACCTATCCGTGGCCCCCTGATTTCTGATCAATATGAGAGAAATGATTTGCCAG 2999

OY	448	GLYGLYTHRProValCYSCysValaArgCysValaPrometGlyLysGlyLIleGluThrLys	467
Db	3000	GGTGAATCCCGCTGTTGCTGCTGTTGTGTAACCTAATGGGAAGGGATTGAACCTAAA	3055
OY	468	Leu-----	468
Db	3060	CTGTGAGTAACTAATTCTTAGACCTTAACCCCTTAGAGTTTCTTTTTCAGAGGATTT	3119
OY	468	-----	468
Db	3120	ATATATTTTCATTTTCATTTCTGATGGAAGTAACTATATAGATAGTACATTTTATT	3175
OY	468	-----	468
Db	3180	TTACTATTACTTGTGTTAGTTCTTGAGANGTCTTGATTTTTCATGTTGATTCATTTT	3235
OY	469	-----ProGluVal	471
Db	3240	GCATTGGCCCTCAATTACTGACTGTTGTTTTTTTAAATAATGATTATAGCCCTGAGGT	3295
OY	471	lValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysIleValAlaAlI	491
Db	3300	CGTCATATGTTCCAGCAAAAGACTGATTAATATGCTCCCTAAACATTAACGCCGTTGCAT	3355
OY	491	egLYCysGlyIleValPhePheValProAspArgGluGlnAspPheAlaSerTYrThrGly	511
Db	3360	TGGATGTAGATCGTTTTTTTTCGTAACGACAGGGGAAGAAGATTTGGTCTTAACAGTA	3415
OY	511	uPheLeuArgTYrLeuSerSerIleAspArgAlaGlyValAlaLysLeuAspAspGlyTYH	531
Db	3420	ATTTCCTCGGTAACCTTACTCAAAAGATCGGGCGGGTGTGCCAAATTAGATGATGTAAC	3475
OY	531	rThreLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGly	551
Db	3480	AACTTATATCTTGGTGCTCCATCAGATTTCTTAATCATGATGTACTCAAGTGAACCGCTCA	3535
OY	551	ngLuArgLeuTYrGlyValValIleuLYLeuProProProAlaValProValThrAlaSer	571
Db	3540	AGAACGGGTAAATGATGTGTGTTCTCAAGTTACCCCGGCAGCGCTTCTGTTACAGCATTC	3595
OY	571	rTYrArgGlnGluSerGlnSerAsnProLeuHisTYrMetAspGlnAlaArgAspSerPr	591
Db	3600	ATACGACAAAGATTCAGTCCAACTCTGCAATTATATGATGATCAAGCCCGGATTCACC	3655
OY	591	oAlaAsnAlaSerHisSerIleuTYrProProArgGluAsnTYrIleArgGlyAlaProGly	611
Db	3660	TGCCATCTGATGTCACAGTTATATACCTCCTAAGGAAATTAACATTAGGGGTGACACAGA	3719
OY	611	uHisLeuThrAlaAlaSerLYsProSerValSerGluProLeuArgGlyLeuProAsnAla	631
Db	3720	ACATTGTGACACTGCTTCAAAAACCATCTGTTAGACGAGCTCTCAGAAATTAATAATGAC	3779
OY	631	aAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLe	651
Db	3780	AGCGCTCAAGCTGGGGTTAGTTTAACCTCCGAGGTTTTAGCACTTCGGCAATCTATCT	3839
OY	651	uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh	671
Db	3840	CCCTGCAACTTCTCAACTGCTGCCCTCGAAGATCAACCACTTAATGTACAGACCTTCAAC	3895
OY	671	rValValSerThrAlaHisGlnSerAsnGlyLYeunTYrAsnGlyGlyAlaProSerGlnAl	691
Db	3900	AGTTGTTTCCACAGACATCACTGCTCAATGAGACTGTACATGAGAGAACCGCTTCAAGC	3955
OY	691	aTrpLYsArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTYrGly	711
Db	3960	TTGGAAAAAGAGTCCAAACAGTTTCATGATGCGTCAATCAATCATTCACAAATACGG	4019
OY	711	LYsAsnGlnTYrThrProAlaGlyGlnLeuProProProProSerArgTYrProProAlaSe	731
Db	4020	AAATCAGTACACTCGAGCTGGGCAACTACCTCTCTCTTCGGGTTACCTCCACAGCTTC	4079
OY	731	rAsnAsnProAsnTYrThrSerGlyMetValHisGlyAsnMetGlnTYrGlnSerGlnSe	751

Dd		4080	AAACCAACCCCACTACACTAGTGGATGTCATCGCAACTGCATTGCCAGGCCAATC	413
Oy		751	rYlaaenMetProGlnLeuSerPerProleuProAsmMeProHisAsnAntyrSermetTy	777
Dd		4140	TGTTAACAATGCCCTGAGCTGTCTCTCCGTTACCAAATAATGCTCATATATATTTCATGTA	4156
Oy		771	rThrGInglySerSerAsnHisProValSerGlnProMetValGlnIntyrGlnProGI	799
Dd		4200	CACCTCAGGGGTTCGTCAATCATCTCGTTTCTCAGCCCATGGTCCAGCAATCCCAACAGA	4225
Oy		791	uAlasarmetProAsnGlnAsnTyrglyProIleproserTyrgIngnlalaasnPheH	811
Dd		4260	AGGCTCCATGCCCCAACCAAACTAATGTCATTCCAATTATCAGCAAGCTAATTTTCA	4333
Oy		811	sGlValaThrThrAsnglnalaglnAsnLeuAsnProSerGlnPheglnalalaamEtG	833
Dd		4320	TGGCGTACACAACAAANTCAGGCCACAAACTTAAACCCTTCCCANAATTCAAGTGCACATGCA	4373
Oy		831	nProProAlaaplyValaAsnLeuglnProGlnAsnglnalaleuArgLeuglnProMe	855
Dd		4380	ACCACACAGACAGATTAAGGCAAAATTITTAGAGCCACAAAACCAACACTACATATGCAGCCTAT	4433
Oy		851	tLieserGlyAsnPGlyGInglyThraThrAsnGlygluValaspIysAsnglnargTyrgI	877
Dd		4440	GATCTCTGGGGAGTGTCAGGGGTACACAGATGGGGAGGTGCATATGAATCAGAGATTCCA	4493
Oy		871	nSerThrLeuglnPhealalaAsnLeuLeuglnlleGlnInlysgIngnInglnI	891
Dd		4500	GTCAAACATRCNAATTTTGACGCAAACTCTTCTCCAGTACAGCAGAAACGACGCAACA	4555
Oy		891	nSerSerGlyThrProalaglyGInglyPro	901
Dd		4560	GTCCTCAGGTACTCCGGCTGCAAGGGGCT	4590

RESULT 3

```

US-09-920-705-4/c
; Sequence 4, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz D.
; APPLICANT: Michaels, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296, 97214
; CURRENT APPLICATION NUMBER: US/09/920,705
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: FPA antisense fragment
US-09-920-705-4
    
```

Alignment Scores:

Pred. No.:	6,03e-121	Length:	801
Score:	1436.00	Matches:	265
Percent Similarity:	99.62%	Conservative:	0
Best Local Similarity:	99.62%	Mismatches:	1
Query Match:	29.91%	Indels:	0
DB:	10	Gaps:	0

US-09-920-705-3 (1-901) x US-09-920-705-4 (1-801)

Oy	197	AspPheLyGlyASPValGlnProserLyValIleutRplleGlyPheProProthrAla	216
Dd	799	GACTTTAAAGAGAGATCTCCAGCAGATGAAGTTCTGTGGATTTGGCTTCCCTCTACAGCT	740
Oy	217	ThrGInCyAsnaSPgluGlnIleuuhIsanaIawetiIeuNeuPeglyGluIleGlu	236

```
Db 739 ACACAATGCAATGATGACAAATTCTGCACATGCAATGATCTCTTGGTGATCGAG 680
Qy 237 ArgVallySerSerYrProSerArgAsnPhaAlaValGluPheArgSerAlaGluGlu 256
Db 679 AGGGTAAAGATTACCATCAGAGAAATTTTGCACTTGAGATTGAGACGGGAGAA 620
Qy 257 AlaArgGlnCysIleValGluGlnArgLeuPheAsnProArgIleVal 276
Db 619 GCTCGCAATGCAAGAGAGCCCTACAGGAGGATTTCAATTAATCTTAGAATCAAAATT 560
Qy 277 MetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMet 296
Db 559 ATGTACTCAACAGCATGAGTTGCTCCCTGACAGCAAGCATCTAGTTTATCTGTGTAAG 500
Qy 297 LysArgSerArgThrAspMetPheAsnAspProSerCysValSerProHisSer 316
Db 499 AAACGGCAGAGACAGATATGTTCAATATGATCTTCATTTGTATCTCTCTCATTTCT 440
Qy 317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
Db 439 ACTGGAAATTCCTGGGTCTATGAGGCCCTCAGAGGTAACAATGAGCCTTCATTAATGGT 380
Qy 337 AlaGluTyrAsnAspValValGlyGluProAsnTyrArgArgProSerAlaAsnGly 356
Db 379 GCAGAAATACAAATGACGTTGTGGTAAAGAGCCAAACTGGAGAGGCCATCTGCAATGGA 320
Qy 357 ThrGlyIleLeuProSerProThrArgIleProGlyIleLeuProSerProAlaGlnGlyThr 376
Db 319 ACTGGAATACTCCATCTCCAAAGAGACCTGGAAATCTCCATCTCTGACAAAGGTAGG 260
Qy 377 ArgArgProMetArgSerAsnProAspSerTyrGluGlyTyrAspProAlaGlnLeuVal 396
Db 259 AGCGCGCTATGAGGTCAAAACCCGATTCCTGGAGAGATATGATCTCTCATGTTGGTC 200
Qy 397 ArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyVal 416
Db 199 AGGAAGATMAAGAACCAAGACAGATGATCAGTACGACGTTTACTCCATGAGGTGTC 140
Qy 417 AspGluArgSerPheGlyArgIleArgIleAlaAlaArgProIleArgGlyProProAsp 436
Db 139 GATGAGAGTCAATTTGGTGCAGGTTCAAGTTGCTGCTAGACCTATCCGTGACCCCTGAT 80
Qy 437 SerAspHisIleTyrArgIleMetIleAlaIleValGlyThrProValCysCysAlaArg 456
Db 79 TCTGATCAACATATGAGAGGAATGATGCAAGAGTGAACCTCCGTGTTGTGTCTGT 20
Qy 457 CysValProMetGlyLys 462
Db 19 TGTGTACTATGGGAAAG 2

RESULT 4
US-10-425-115-123950
; Sequence 123950, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 123950
; LENGTH: 3463
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_44523C.1
US-10-425-115-123950
```

```
Alignment Scores:
Pred. No.: 1,666-106
Score: 1286.50
Percent Similarity: 49.80%
Best Local Similarity: 35.63%
Query Match: 25,80%
DB: 18 Gaps: 40

US-09-920-705-3 (1-901) x US-10-425-115-123950 (1-3463)
Qy 16 GlnSerAsnAsnLeuTyrValGlySerLeuThrProGluThrGlnLysArgLeuThr 35
Db 227 GAAAGCAACAGCTCTGGGTGGGCAACCTGCCCTCGACAGTACCCAGAGGAGATTGTG 286
Qy 36 GluLeuPheGlyArgTyrGlyAspIleAsp--ArgIleThrValTyrSerArgGly 54
Db 287 GCGCTGTTCCGGCCGCGCGCGCGCTCGATTGCCCTTGACAGAGCCGATCTCGCAGC 346
Qy 55 PheAlaPheIleTyrTyrArgHisValGluGlnAlaValAlaIleValGluGln 74
Db 347 TAGCGGTGTCCTCTTCCGTTCCCGACCGAGGCTCGCGAGCCGTCGAGGACCCGG 406
Qy 75 GlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCys 94
Db 407 GGTGAAAGGTCAAGGGGGCGCCATGGCACCCAGATTCGGCGCGCGCTAGAGCTGTT 466
Qy 95 LysSerLeuTyrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGlu 114
Db 467 AGAAATCTGTGGGTGGGATTAAGCCATTCATTCMAAGAGAGAGCTTGAGGGGAG 526
Qy 115 PheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPhe 134
Db 527 TTTCTAAAGTTCCGAAAGTTGAGAGTTCGTTCTCCCAAGATCAACTTCACGATAC 586
Qy 135 IleAspTyrTyrGluMetAspAspAlaLeuGlnAla--LysSerMetAsnGlyLysPro 153
Db 587 ATAGATTTTGAAGAGCTTGAAGACGCAATTTCTGCTCAAGATCTTGAATGAAAAATG 646
Qy 154 MetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGluGln 173
Db 647 CTAGGTGGCAAGAAATGTGTGATTTCCAGAGGTCCAGAGAGCGAAGCGAATGTGCA 706
Qy 174 TrpAlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHis 193
Db 707 GAGCTTACAGCTTCAAGTGA-----GTACCAAGG 739
Qy 194 SerTyrGluAspPheLysGly----- 200
Db 740 CCAGTGGGTGATTAAGAGAGGAGGAGTGCCTCTGAAGATTCTGCTGAGTGGATGCGA 799
Qy 201 AspValGlnProSerLysValLeuTyrIleGlyPheProThrAlaThrGlnCysAsn 220
Db 800 GAGGCAACAACCTACCAATGTTCTCTGGGTGTTT---CCAGGTTCTTATGAGTCATT 856
Qy 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGlnIleGluArgValLysSer 240
Db 857 GATGAGAGGACACTTAGCATGCAATGCTATTCGTGTCCTTAGAGATTAAGATT 916
Qy 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlnAlaArgGlnCys 260
Db 917 TTCCAGACAAAGCAGTATGCTTTGTTGATTTGCAATGTTGGGAAAGCTTGAATGCT 976
Qy 261 LysGluGluLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db 977 AAGATGAATCTAATGTGATCTTTTCAATGATCCAGATTCGATTCGATTCCTCAAC 1036
Qy 281 AspGluLeuProProGluGln--AspAspThrSerPheTyrSerGlyMetLysArgSer 299
Db 1037 AGTGGGCTTGACCAACAAATTTGACACCAACCAATCGGTGCTGATTTCTTCAATGCA 1096
Qy 300 ArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIle 319
```

Db 1097 -----GATATCTATTCAGTGAATGCTGTAGAGCCCTGTATTCAGTGAACCTTTG 1150
 Qy 320 ProGlySerMetArgProLeuArgGlyThrAsnGlnArgSerTyrAsnGlyAlaGlyTyr 339
 Db 1151 CAAGGG---TATATTCACCAAGAGAGAGAGAACAAAGA-----TAT 1189
 Qy 340 AsnAspValAlaGlyIleGlyProAsnTyrArgArgProSerAlaAsnGlyThrGlyIle 359
 Db 1190 TAGGATTTATGCGGGT----- 1204
 Qy 360 LeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgPro 379
 Db 1205 -----ATGCCCTACTCCAGGTGGATCTTTCCCAACT----- 1237
 Qy 380 MetArgSerAsnProAspSerTyrGluGlyTyrAspProAlaGlnLeuValArgGlySer 399
 Db 1238 -----GAACCATTTGATCCA-----AAGAAGCA 1261
 Qy 400 LysArgThrArg-----ArgAspGlySerValAsp 409
 Db 1262 AAAAGATAGAGCTGTGCTGTGCTGACCTTACATGTAAAGGAGGACAGCAGAGT 1321
 Qy 410 -----GlyPheThrPrometGlyValAspGluArgSerPheGlyArgGlySer 425
 Db 1322 CTTTATTCCTGCGATACGTCATCAGGGCAGTTCTGTGCAAGCTGAGGAGAGCTCAAGT 1381
 Qy 426 ValAlaAlaArgProIleArgGlyProProAspSer-----AspHisIleTyr 441
 Db 1382 CCGTATATTGCA---GTCCGGGGCAGAGTCATCATCATCATCATCATCATCATCATCAT 1438
 Qy 442 ArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValPrometGly 461
 Db 1439 CGTGGCAGATTCGCAAAAGGTGATCCCTGTTGTGCTGCTGCTGCTGCTGCTGCTGCT 1498
 Qy 462 LysGlyIleGluThrLysLeuProGluValAlaAsnCysSerAlaArgThrAspLeuAsn 481
 Db 1499 AAGGGCAGTGAATACCTTACCGAGTGTATTAATTTGCTCAGCTCGAGCTGATTTGGAT 1558
 Qy 482 MetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAsp 501
 Db 1559 ATGCTGCGAAGACATACCGGAGTGCATCTGATTTGACATTTGCTTTTCTTTCGCAAGAT 1618
 Qy 502 ArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArg 521
 Db 1619 AGCGAAGATGACTTGTGTTCTTATCTAGAGTCTTGCTGCTTACCTGGGCTCAAAAGTCGG 1678
 Qy 522 AlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPhe 541
 Db 1679 GCAGGGGTGTAAAGTGTATGACAGGAGCTACATTAATTTTGTTCACCATCAGATTTTC 1738
 Qy 542 LeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeu 561
 Db 1739 CTAAACAATGTGTAAAGTGTATGATGTCACAGCGGCTTATGTGTGTATTAACATTT 1798
 Qy 562 ProProProAlaValProValThrAlaSerTyrArgGlnGlu-----SerGlnSer 578
 Db 1799 -----CCACAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1852
 Qy 579 AsnProLeuHisTyrMetAspGlnAlaArgAspSerPro---AlaAsnAlaSerHisSer 597
 Db 1853 CAACCT---TACTATGATGAAGGAGGAACTCTGCTTACACAAAGAAAGTATAGT 1906
 Qy 598 LeuTyrProProArgGluAsnTyrTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSer 617
 Db 1907 ATCATTTCTCTATATACAGT-----GGCCACCTTGATGTGATAT 1948
 Qy 618 LysProSerValSerGlu-----ProLeuArg----- 626
 Db 1949 CGGACATCTTTTGCATAGGATTAATGATCGGTTGGGGCATATTTCTTGAAAGGCTCGG 2008
 Qy 627 -----IleProAsnAsnAlaAlaPro 633
 Db 2009 GTGATGAAGGCCAAGCAGTTCAACAGCTCTTGAGGTTTCTCTATATCAAGCAGCT 2068

Qy 634 -----GlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAla-SerIleLe 651
 Db 2069 GGGCTGCAATCAATCTTCACTCAAGCTCAATATATGCTTACATTTGGCAAAACCTTCT 2128
 Qy 651 upProAlaThrSerGlnProAlaAlaProGluSerHisGlnPrometSerGlyProSerThr 671
 Db 2129 GCTTAGTGTG-----CAATGTTCACCGCTGTAGTGTAGTGTAGTGTAGTGTAGT 2170
 Qy 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAl 691
 Db 2171 TCCCATAGACAGAACATCGCAG-----ATACGAGATCATCATCATCTTTCAAAAGT 2221
 Qy 691 atPrrLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlyTyrG 711
 Db 2222 ATGG-----AATCCGAAACCAAGTATCTGCTTCAAAATTAATCTTTGGGCAAAATGGC 2275
 Qy 711 yAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgTyrProProAlaSe 731
 Db 2276 TAAATGTTCAGATCCAGACAGCAGATTT-----ATGGCCAAAGCTTCAGCTGC 2323
 Qy 731 rAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMet---GlnTyrGlnSerG 750
 Db 2324 TCATTTGACAAACTAT-----GGAAACATGAGTGTGAGTGTGAGTGTGAGTGTGAG 2365
 Qy 750 nSerVal-----AsnMetProGlnLeuSerPr 759
 Db 2366 CTCATATCAACAATCTCTTACACCTCGAAGTACTTTGAATCTTGCCCAACCACTTC 2425
 Qy 759 oLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisPr 779
 Db 2426 ACTTCCAAACAATCAACACAGCTTGTCTACATTAAGTCTCAAGTCTCAAGTCTCAAGTCT 2485
 Qy 779 oValSerGlnPrometValGlnGlnTyrGlnProGluAla---SerMetProAsnGlnAs 798
 Db 2486 C---TCACAGACAAACAGCACTGATATCAGCAGAGCAGATTAATGTCGCCCAAAACA 2542
 Qy 798 nTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGln-- 817
 Db 2543 CTATGGTCCATTATGTTCCAGTTAGGCCATTTCACTTCAATCAGCAACACTTACAAACC 2602
 Qy 818 -----AlaGlnAsnLeuAsnProSerGlnPheGlnAlaIleMetGlnProProAl 834
 Db 2603 CACCTTACCATCCCAACAAGTAACCTTGA-----CCTCCAC 2641
 Qy 834 aAspLys-----AlaAsnLeu---GluProGlnAsn-----GlnAlaLeuArg 847
 Db 2642 AATATATCAGATTGGGAATTTGGCCAGCAGACATCTATGCCACATGATGATGATG 2701
 Qy 847 gLeuGlnPrometIleSerGlyAspGlyGlnGlyThrThrAspGly----- 862
 Db 2702 ACCAAGTCAGATTTCTCTTCAAGGGGCACAAACAAATCGTGTCTGTGTGTGTGTGTGT 2761
 Qy 863 -----GluValaAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLe 880
 Db 2762 AGCTCCCAAGAGATTAAGCAAGAGTACAGGACATCTCCAGTTAGCTCAAAACCT 2821
 Qy 880 uLeuLeuGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 893
 Db 2822 ACTGCTTCAAGTACAGCAGCGTGAATCGAATAATCATCTGAGACCTAATTAATGCTTC 2881
 Qy 893 rGlyThrProAlaGlyGlnGly 900
 Db 2882 AGGTACCTCGTCTGTGTGTGT 2903

RESULT 5

US-10-437-963-36321
 ; Sequence 36321, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua

Alignment Scores:		
pred. No.:	4 06e-106	3084
Score:	1281.50	339
Percent Similarity:	51.4%	Conservative: 150
Best Local Similarity:	35.6%	Mismatches: 306
Query Match:	26.6%	Indels: 155
DB:	18	Gaps: 38

QY	223	PheGlyGluIleGluArgValIlySerTyrProSerArgAsnPheAlaLeuValGluPhe	251
Db	805	CATGGTGGCTGTACAAATACAAAACTTTCCCAACAGGCAATATGCTTTTGAGGTTT	864
QY	252	ArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsn	271
Db	865	GCAACAGTTGGGGAAGCTCTAATCTAAGAAATCTAAGATGTCGCTTTTCAATGAT	924
QY	272	ProAqGlyLeuValIleMetTyrSerAsnAspGluLeuProProGluGln---AspAspThr	290
Db	925	CAGAAATTCAAATCTTTCTTCGAAACGTGAGCTTGCGCCAAATAAATTATGATATCA	984
QY	291	SerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspProSerCys	310
Db	995	ACAGAGATTTCTGGATTTCCAAATCA-----GAGATGATATTAGATAGTGTCAATAT	1033
QY	311	ValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsn	330
Db	1039	GGCGCTTCTGATTAT-----TTTGATCTTCGACGGGGAAGATCA	1077
QY	331	GluArgSerTyrAsnGlyAlaGluThrAsnAspValAlaGlyLysGluProAsnThrArg	350
Db	1078	AGATTC-----TTTGAGTACAGTGCTGC-----	1107
QY	351	ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPro	370
Db	1102	-----CCTGTTCTTCGGTGTATTTCTTCA	1122
QY	371	SerProAlaGlnGlyThrArgArgProMetLysSerAsnProAspSerTrpGluGlyTyr	390
Db	1126	TCACCGAATCTGA-----AATCTCTCTTAATCTGACAGCATCTGCACAGTACATTC	1179
QY	391	AspProAlaGlnLeuValArgLysSerLysArgThrArgArgAspGlySerValAsp---	409
Db	1180	GATCA-----AGAGAAGCCAAAGCTTGCGGTGATGCTGCAGCTGACCT	1222
QY	410	-----GlyPheThrProMetGlyValAspGluArgSerPhe	421
Db	1228	TATGATACAAAGGAGAGTAGTAGGAGGCTTATTCTGCTGGGTACTTCACGCGCAGAGT	1287
QY	422	GlyArgGlySerValAlaAlaArgPro-----IleArgGlyProProAspSer---	437
Db	1288	GCAAGTCTTGAGAGAGAGCTCAAGTCTCTGATTGGAATTCATGGCACAGTGCACCGCA	1347
QY	438	-----AspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys	454
Db	1348	TCATACCTTGAGCACTTCTGGCGGTGATGATGCCAAAGGTGATGCTCTCTGTGAT	1407
QY	455	AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys	474
Db	1408	CTTACACATATTCACATG-----TTTGCAGACCGAGATGTTGTAATGTT	1452
QY	475	SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu	494
Db	1453	TCCGCTGAAGTGAAGTATGATCTGGCAAGCATTTAGAGACGCTTCAGGGTTTGAT	1512
QY	495	IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg	514
Db	1513	ATGCTCTCTTCTTGGCCAGATAGTGAAGAGACCTTGTCTTACCTGATTTTTCGCG	1572
QY	515	TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe	534
Db	1573	TACTTAGCTCAAAAGATCGGGCAGGGGCTTGTAAGGTGATGGGGGACTACTTTATTT	1632
QY	535	LeuValProProSerAspPheLeuThrAspValIleGlnValThrArgGlnGluArgLeu	554
Db	1633	TTGGTCCACACTTCGATTTTAAAGAAAGTTTTCAGAAATGTTGCAATGATGCTCCAGAAC	1692
QY	555	TyrGlyValValLeuLysLeuPro-----ProProAlaValProVal	568
Db	1693	TACGGTGTAGTATTACACATTTCCGCAAAATGTCTGCTGCTGCTCTCTGCTCCACT	1752
QY	569	ThrAlaSerTyrArgGlnGlu-----SerGlnSerAsnProLeuHisTyrMetAsp	585

```

Db      1753 CCTGCTGTCAAAAGGCACACACTAAGTGCACGAGATTCACAACCT-----TACTAGAGT 1806
Qy      586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuThrProProArgGluAsnIyr 605
Db      1807 GAA---AGGGAATTCCTTGTGCAGAGAGAGTATGATGATTACGGCTTAGCAACAATTCAC 1863
Qy      606 IleArgGlyAlaProGluHisLeuThrAla----- 615
Db      1864 CATCGAGATGCT--GATCACCGGTGATCTTACGTGAGAGATTCAATTCATCAATTGGGG 1920
Qy      616 -----AlaSerIyrProSerValSerGluPro-----LeuArg 626
Db      1921 CAATTCCTTGACAGACTCGGGGTGATGAAGACAACTAGTTCAGCCAAACCTTGGCCGT 1980
Qy      627 IleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuAlaThr 646
Db      1981 ATCCCTTACAAACCCAGAGACTACAGTACAGCCTTCACTCCAGCTGATATGATAGACACT 2040
Qy      647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMet 666
Db      2041 TTAGCTAAACTTTTGGCCAGGTGGGAGTCATCAGCTCTGTGTCATCGCCAGCTTCCCTG 2100
Qy      667 SerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuIyrAsnGlyGlu 686
Db      2101 AGC-----TCACAGACCAACCACTGCATTAACCCAAATGAT-----GATGCATCT 2145
Qy      687 AlaProSerGlnAlaThrIyrAsnArgGlyProGlnThrValHisAspAlaSerAsnGlnSer 706
Db      2146 ACTCTTCAAAAGTATAGAGAG-----CTGAAATCAAGCAATGAGCTTCTACTTCATCC 2199
Qy      707 PheGlnGlnIyrGlyAsnGlnIyrThrProAlaGlyGlnLeuProProProSerArg 726
Db      2200 TTAGAGCAAGTCCGAT--TTCACGACATTCAGACAA-----CAGTTCAACA 2247
Qy      727 TyrProProAlaSerAsnAsnProAsnIyrThrSerGlyMetValHisGlyAsnMetGln 746
Db      2248 CAAGCTGGAGCACTTATCTGCCAAACTAT-----GGGAAGCTTGGCAGGTGCACAGAG 2301
Qy      747 TyrGlnSerGln-----SerValAsnMetProGlnLeuSer 758
Db      2302 CACCCCAACCAACATAGTGTCTTACAATCTCGAATAGATCTTAATCACTTGCACACACCA 2361
Qy      759 ProLeuProAsnMetProHisAsnAsnIyrSerMetIyrThrGlnIyrSerSerAsnHis 778
Db      2362 CCACTCTTACGCTGCACCCAGCTTCTGTATATTATTCATCGCAAGTTGCACACAGTTTG 2421
Qy      779 ProValSerGlnProMetValGlnGlnIyrGlnProGluAla---SerMetProAsnGln 797
Db      2422 CCT--ACGCAAGTAGTACAGCAACGATATCAACGACAGCATATTATATGACTCAAGC 2478
Qy      798 AsnIyrGlyProIleProSerIyrGlnGlnAlaAsnPheHisGly----- 812
Db      2479 AATTATGATCAATTGGCACAGTATAGATCTTAATCTTACAGCGCACCATCAACAAT 2538
Qy      813 ValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnPro 832
Db      2539 GTTGCACTTCCTGCTCAGCAAGCTCTGTGCTGCAGCAATTCCTCTGCTATGACAGCT 2598
Qy      833 ProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIle 852
Db      2599 CTGTCT-----GCCGACAAAGCTCCTGTT 2622
Qy      853 SerGlyAspGlyGlnIyrThrThrAspGlyGluValAspLysAsnGlnIyrGlnSer 872
Db      2623 GCT-----GCACAGAGTCTGTCTGAC--GAGGCAAGAGCAAGAAAGATTCAGGCG 2673
Qy      873 ThrLeuGlnPheHisAlaAsnLeuLeuGlnHisGlnGlnIyrGlnGlnGlnIyrGlnSer 892
Db      2674 ACTCTTCAAGTGTGCTCAGCGCTGTGCTGCTACACTCAACAGAAAGCTGGAATCAACCT 2733
Qy      893 SerGly-----ThrProAlaGly 898

```

```

Db      2734 TAAGGCTATCTTGAAATCAACCTTCAGGC 2763
RESULT 6
US-10-437-963-79215
; Sequence 79215, Application US/10437963
; Publication No. US20040123343A1
GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 79215
; LENGTH: 3315
; TYPE: DNA
; ORGANISM: Oryza sativa
FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_78947C.1
; US-10-437-963-79215
Alignment Scores:
Pred. No.: 1,9e-62 Length: 3315
Score: 802.00 Matches: 288
Percent Similarity: 37.298 Conservative: 120
Best Local Similarity: 26.584 Mismatches: 325
Query Match: 16,704 Indels: 378
DB: 18 Gaps: 45
US-09-920-705-3 (1-901) x US-10-437-963-79215 (1-3315)
Qy      7 ProPheArgAlaAspAspSerGlyPhe-----GlnSerAsnAsnLeuTrpVal 22
Db      121 CCGCGCGGCGGCTCTCTCCGCGGTGGGGGTGCACCGCCGTCGCGGCACTGTGGGTG 180
Qy      23 GlySerLeuThrProGluThrThrGluSerAspLeuThrGlnLeuPheGlyArgIyrGly 42
Db      181 GCGAGCTCTCCCGCGCGCTCCCGCGCGCGGCGGCACTTCCGAGCTCTTCCCGCGTGGC 240
Qy      43 AspIleAspArgIleThrValIyrSerSerArgGlyPheAlaPheIleIyrIyrArgHis 62
Db      241 GACGTGAGAGGCACTTCCCGGTGACCCCGGCGGAGCTTGGCTTGTGACGTTCCGCGCG 300
Qy      63 ValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyValAsnLeuAsnGlySerGln 82
Db      301 GAGGAGGAGCGCGTGGCGCGCGGTGGCGGAGCTGCAGGAGATCACTCCGCGCGCGGCC 360
Qy      83 IleLysIleGluIyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyIle 102
Db      361 ATTAGGATGAGATTTTCCAG-----GGGAGTAA 390
Qy      103 GlyProAsnValSerLysAspAspLeuGluGluPheSerLysPheGlyLysIleGlu 122
Db      391 GGTTCAGTAGCTCTTATGATGAC-----AGATCTCAACAACATGCT 432
Qy      123 AspPheArgPheLeuArgGluArgIyrThrAlaPheIleAspIyrIyrGluMetAspAsp 142
Db      433 GATCAAGAGCGTTTACTGACGAGGA----- 459
Qy      143 AlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValaAsp 162
Db      459 ----- 459
Qy      163 PheLeuArgSerGlnAlaProLysGlyGlnIyrPheAlaGlySerIyrAspAsnArgAsn 182

```

Db 460 -----AGGATGACGACATCAAGTCTGAAAA-----TCAGCTGAT 495
 QY 183 GlysMetAsnHisIysProGlnTyrProHisSerTyrGluAspPheIysGluAspVal 202
 Db 496 AAATCCAAAAGAAAGCGCCA-----GCA 519
 QY 203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
 Db 520 GAACCTGATGAGATATGATGATAGGTTTCTGTTGGTCTGAGGTA-----GATGAG 573
 QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
 Db 574 GCAACTCTCTGGGAGAGCTTTTCACTTTGGTGGAGGTGTCAAGATTAACCTACATTCCCA 633
 QY 243 SerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlnAlaArgIleCysIysGlu 262
 Db 634 GGGCGCTACTTATGATTTGTCAGTACACTACTATTGTCAGCGGACATGCGGGGAGAA 693
 QY 263 GlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSer----- 279
 Db 694 ACACTCAGGGAATATTTTTCATTAACCTCGAGTTAGCATTTGCTTTCTCGAGTGCAC 753
 QY 280 -----Asn 280
 Db 754 AGTGTTCAGCAGAAATTTGGAAAAAGTCTTAGATGCCCAATATCCCCCATTTAAAC 813
 QY 281 AspGluLeuProPro-----GluGlnAspAspThrSerPheTyrSerGlyMetLys 297
 Db 814 TCTAGTGTATGACCTTATTCAGGGAGAGATTTTGAAGATTT----- 858
 QY 298 ArgSerArgThrAspMetPheAsnAsnAspPro-----SerCysValSerSerProHis 315
 Db 859 ---CCTAGGGCTAGGCTTTGTAGTCTCTCAAGAGATATGACATGTCATCTCCACAT 915
 QY 316 SerThrGlyIleProGlySerMetLysProLeuAlaGlyThrAsnGluArgSerTyrAsn 335
 Db 916 TAT-----GGCCCTAAGAGACTTCTTGA----- 939
 QY 336 GlyAlaGluTyrAsnAspValAlaGlyLysGluProAsnTrpArgArgProSerAlaAsn 355
 Db 940 -----GATCATGATATATGTGGTTTCAGCAGAGATATATATTGCGA----- 981
 QY 356 GlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerPro----- 372
 Db 982 -----TATGACCTGAGTAGAGCTTGATCTGATCTATAT 1017
 QY 373 -----AlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
 Db 1018 TTGGAACCTTTAGGATACAGGGCTCGGTCCAGAAAGAGATGTGTAGAGACCATAT 1077
 QY 388 GluGlyTyr----- 390
 Db 1078 GAACAGCATAGGCGTAGCCCTGCTGTGATGACCATGACGACAAACATTCCATTGACGGA 1137
 QY 391 -----AspProAlaGlnLeuValArgGlu----- 398
 Db 1138 TCTCAGGAGCCCTTACCATTAAGAGATGTCGGTATGCTAGGAGAGATCCATACCATTT 1197
 QY 399 SerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGlu 418
 Db 1198 TCAGAAAGATTGAGACTGTGTAGAGACATGACTCTGAACTCTTC---GAATACCTTTTC 1254
 QY 419 ArgSerPheGlyArgGlySerValAlaAla-----ArgProIleArgGlyPro 434
 Db 1255 TCTGATTTGATGAGGGGAGAGTGTGCTGCTGCTACCCAGAGAGGCCCTTCTATGTGTG 1314
 QY 435 ProAspSerAsp----- 438
 Db 1315 CCAAGATGATGACATACACCCAGAGGCTATCAACTGTCTCATGATGTAGAAATCAT 1374
 QY 439 -----HisIle----- 440
 Db 1375 GTTGATCTTTTAGGAATCCAACTCCTGTTGATGATGAGCATATACAGAGGACATGACAG 1434

QY 440 ----- 440
 Db 1435 GACAGCTTTCTAGGCAATGATAGAGTGAAGATCACTCTGTAAATACATGACCCCTT 1494
 QY 441 -----TrpArgGlyMetIleAlaIleAlaGlyGlyThrProValCysCys 454
 Db 1495 CTCAGAGGAAGATGAAATGGATGGTATCAATAGCAAGGAGGACACCAATTTGCCCCA 1554
 QY 455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValAlaAsnCys 474
 Db 1555 GCGCATGCTCTCCCTGTTGGAGAGGTCTTAACCTCATGCTGCCGGAATTTTGGATTGC 1614
 QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu 494
 Db 1615 ACTGCTAGAGACAGCTGAGATGCTCTCTTAAGCATATTTTCAAGCTGCCAGCAGCTGC 1674
 QY 495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
 Db 1675 GGGGTGTTTGTTCAGAAATATGCTGTACATGCGACGCTTATGAAATTCATGAT 1734
 QY 515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe 534
 Db 1735 TACCTTGGTGAATAGCAGCGTGCAGAGTTTGTAACCTTGAGAAAGAGCAGCTTATTT 1794
 QY 535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
 Db 1795 CTGTTCACCTCCAGACTTCTTGACACMAAGTACTGAGGTTCCAGGTTAAAGTCAGCATA 1854
 QY 555 TyrGlyValIleLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGln 574
 Db 1855 TGTGAGTCACTTGTGAATTTGAGACAGTCAGATCCAGAAAGTTTCTGCCCACTGCCAAA 1914
 QY 575 GluSer----- 576
 Db 1915 CCAGAAATTTGTAGTCAATTGTAACCATGATGTCGTGCTCATGAGATCTAGATGCA 1974
 QY 577 -----GlnSerAsnPro-----LeuHisTyrMetAspGlnAlaAspArg----- 589
 Db 1975 TTGAGAAATCAACCCACAGATATGAGGCCACTCTCTCAGGGTTCAAGATATTCGGCG 2034
 QY 590 ---SerProAlaAsn-----AlaSerHisSerLeuTyrProProArgGluAsnTyr 605
 Db 2035 TTGTGCGCTGGAACCTTAATCCAGCAAGTSCACATTTGTGTCCGCTTACAG----- 2088
 QY 606 IleArgGlyAlaProGluHisLeuThrAla----- 615
 Db 2089 TTGAAATGCTCTTCATATCTGAATCTGAATTAGCTCATCAAAAGCATCCACTGAC 2148
 QY 616 -----AlaSerLysProSerValSerLysLeuProLeuArgIleProAsn--- 629
 Db 2149 TCCACAGGAGATAGCAGATGACACAGCAGCAGACACCAACCAAGTATATGCTCCCTCAGA 2208
 QY 630 -----AsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 643
 Db 2209 TGTGAGATACATTTAATCAATCCAAATCCAGGTCTGGA-----AATTG 2253
 QY 644 LeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHis 663
 Db 2254 AATTATTTGGCTGAGAGTGGATCCCATATCACTGATAGCAGCAGGAGGACATAC 2313
 QY 664 Gln-----ProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGly 681
 Db 2314 TCATTTCCTCTCAAGAGATACAAA-----GTAGATCA----- 2349
 QY 682 LeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAsp 701
 Db 2350 -----TCAGGAGTATGACCA-----GTTGACATGAG 2376
 QY 702 AlaSerAsnGlnSerPhe-----GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGln 719
 Db 2377 GCATCAAAACATGTCTTACCTCCATGCAACCTGACATCAAGAGGTATGTAACCTCAA 2436

QY	720	LeuProProPro	ProSerArgTyrProPro	729	
Db	2437	CAACSTCCATCTCTCCCACTTATCGCTTCCACAGAGCACTTGACAATTGGCCACTTT	2496		
QY	730	AlaSerAsnAsnPro	AsnTyrThrSerGly	739	
Db	2497	CTTGACACAAACAAACCAACGAGAAAGAGCCCTGGACAGCTTTAAACAAATCCAGA	2556		
QY	740	MetVal	HisGlyAsn	MetGlnTyrGlnSerGlnSerValAsn	753
Db	2557	TTTCATACGAGATTCACATGACGATCTTCATATGATGACACACAGCTCAGGTTCTATC	2613		
QY	754	MetProGlnLeuSerProLeuProAsnMetProHisAsnMetTyrSerMetTyrThrGln	773		
Db	2614	CCCTGTCGCAAACTCATCTTGCCACCTGCTCCG	CCA	2646	
QY	774	GlySerSerAsnHisProValSerGlnProMetValGln	786		
Db	2647	TCTGCATACAGTTACAGATTATGCGCCACACAGTTCAAGTTCAAGTCCACCAATCCA	2706		
QY	787	GlnTyrGlnProGlnAlaSerMetProAsnGlnMetTyrGlyProIleProSer	804		
Db	2707	TCCATTATGATACACCGAATGCTCTGATGCTTCTGCACACACTTTGGCCTTACCCCT	2766		
QY	805	TyrGlnGlnAlaAsnProHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSer	824		
Db	2767	ATGCATCCTTCGGGA	AATCCAGCTCATCTTCCATGCTCTTTG	2808	
QY	825	GlnPheGlnAlaAlaMetGlnProProAlaAspValAsnLeuGlnProGlnAsnGln	844		
Db	2809	AGATCGTTGTCCTCTCTCTTCTGAAAGCCCTCCACCCCTTAGGACGACACATCAAGT	2868		
QY	845	AlaLeuArgLeuGlnProMetIleSerGlyAspGlyGln	Gly	858	
Db	2869	GCTTCACAGGCGCAACTGCGACTTCCCTCTGACACACAGACTAGCCAGCAACGCTCTGCT	2928		
QY	859	ThrThrAspGlyGlnValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla	878		
Db	2929	CAGGAAGCCATCCAGGAGACCTCTCAAAAGCGCTTCAGAGCATTTGGCAGCA	2988		
QY	879	AsnLeuLeuLeuGlnIleGlnGlnIleAsnGlnIleSerGlyThrProAlaGly	898		
Db	2989	ACCTTACTTCAGCAGATTCAGCAACAACTTAA	CCTGCTGGC	3030	
QY	899	Gln	899		
Db	3031	CAG	3033		
RESULT 7	US-10-424-599-59704	US-10-424-599-59704	Application US/10424599		
	Sequence 59704	Publication No. US20040031072A1			
	GENERAL INFORMATION:				
	APPLICANT: La Rosa Thomas J				
	APPLICANT: Kovalic David K				
	APPLICANT: Zhou Yihua				
	APPLICANT: Cao Yongwei				
	TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With				
	TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement				
	FILE REFERENCE: 38-21(53223)B				
	CURRENT APPLICATION NUMBER: US/10/424,599				
	CURRENT FILING DATE: 2003-04-28				
	NUMBER OF SEQ ID NOS: 285684				
	SEQ ID NO 59704				
	LENGTH: 1101				
	TYPE: DNA				
	ORGANISM: Glycine max				
	FEATURE:				
	OTHER INFORMATION: Clone ID: PAT_MRT847_24924C.1				
	US-10-424-599-59704				
Alignment Scores:	5.64e-51	Length:	1101		

Score:	669.00	Matches:	169
Best Local Similarity:	54.57%	Conservative:	40
Best Global Similarity:	44.13%	Mismatches:	77
Query Match:	13.93%	Indels:	97
DB:	17	Gaps:	15
US-09-920-705-3 (1-901) x US-10-424-599-59704 (1-1101)			
QY 282 GluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysAspSerArgThr	3010		
DB 54 CAGTTAACCACTT-----AGTGGATATT-----	74		
QY 302 AspMetPheAsnAspProSerCysValSerSerProHisSerThrGlyLeuProGly	3212		
DB 75 --ATGGAGACAAACATACCAATGCGACTTTTGCTCCACAGT--GGTGT-----	1222		
QY 322 SerMetArgProLeuAspArgGlyThrArgGlnuArgSerTyrAspGlyValGlnuTyrAsnAsp	3414		
DB 123 SerMetArgProLeuAspArgGlyThrArgGlnuArgSerTyrAspGlyValGlnuTyrAsnAsp	1555		
QY 342 Val-----ValGlyLeuGluProAsnTyr	349		
DB 156 ATTATATGGCTCCACAAATTCAGAGTGTACTCCAAAGATGACATGGGTCCAAACTGG	215		
QY 350 ArgArgProSerAlaAsnGlyThrArgGlyLeuLeuProSerProThrGlyProGlyLeu	365		
DB 216 AAAAG-----CCATCTCTCTCCACACGAGGATGCTT	248		
QY 370 ProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyrGluGly	3859		
DB 249 TCTTCTCTGCAACGGGTGTAGGCTTCTTACAAATGCAACTCGGGTGCATGGAGTGTG	308		
QY 390 TyrAsp-----ProAlaGlnLeuVal-----ArgGluSerLysArgTh	4020		
DB 309 CTTCGACATAAACCTTATCATTCAGAGAGACACTTGCTGTCCAAAGATTTCTAAAGCTTC	368		
QY 402 TArgArgAspGlySerVal-----AspGlyPheThrProMet-----	414		
DB 369 AAGATATGATGGACCTTGCTGTGATGAAGGCCCATTTCTTGAAGAAATATGATGA	428		
QY 415 -----GlyValAspGluArgSerPheGlyValArgGlySe	425		
DB 429 TCGGGGTGTAAGCTCTAGAGCAGACATATGGAGTTGACCCAGCTATGATGAGAGTGTTC	488		
QY 425 T-----ValAlaAlaProI	431		
DB 489 TGTTCATATGTAAACATTCAAGGAAGAGTCACTTGCTCCAGTTATGCTCAAGGATTAC	548		
QY 431 eArgGly-----ProProAspSerAspHisIleTyrArgGlyMetIleAl	446		
DB 549 AGCTGGAGTACATGTGTGTGCCCAACCTGATATGTATGATCATATTTGGCGTGGAGATTATGC	608		
QY 446 AlYsGlyGlyThrProValCysGlyAlaArgCysValProMetGlyLysGlyIleGluTh	466		
DB 609 AAAAGGGGGAACCTCTGTGTTTCCGTGCTAGAGTGTACCAATYGGGAAAGGAAATTGGAC	668		
QY 466 rLysLeuProGluValValAsnCysSerLysAlaThrAspLeuAsnMetLeuAlaLysH	486		
DB 669 TGAGCTCTCTGATGTGTGTGATGTCTCACTAGAGAGGAGATTGGAGATATCTCAAAAACA	728		
QY 486 sTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAspH	506		
DB 729 TTATCTGTATGGAATTGGTTTGTGATATGTTTCTTCTTCTGCTGATGATGAGACGATTT	788		
QY 506 eAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLys	526		
DB 789 TGCTTCATATACCGAGTTCCTTCCGTATCTTGTGTGCGAAAAATCGGTGTGTGCCAA	848		
QY 526 sLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspValle	546		
DB 849 ATTGTGTATTAACACACACTTATTTTGTGTGCTCTCTTGTGATTTTCTCAAGAGATTTT	908		
QY 546 uGlnValThrArgGlnGluArgLeuTyrGlyValValleuValLeuProProAlaVala	566		


```
Db 909 GAAAGTACTGGAGCTGAGGCTGTATGGTGCTTTAAGTTCCACCA-----GT 962
Qy 566 lProvalThraLaserTyraGingluSerGlnSerAsn-----ProleuhisTywE 584
Db 963 GCCAAGTAGTGCACCTTATGCAACCAACCATCATTTCCGTGACCAACTACTAGTATAT 1022
Qy 584 tAspGAlaAraGspSerProAlaAenAlaSerHisSerleuTyProProArgGluAs 604
Db 1023 GCACAG-----AATCTCTTTCGACACTGAATATGTTGATTCCTGTAAAGAGA 1076
Qy 604 nTyTlle 606
Db 1077 ACATATT 1083

RESULT 8
US-09-920-705-6
; Sequence 6, Application US/09920705
; Publication No. US2003079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaels, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296.97214
; CURRENT APPLICATION NUMBER: US/09/920,705
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 3715
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: FPA promoter plus intron
US-09-920-705-6

Alignment Scores:
Pred. No.: 1,09e-48 length: 3715
Score: 651.50 Matches: 171
Percent Similarity: 27.23% Conservative: 0
Best Local Similarity: 27.23% Mismatches: 0
Query Match: 13.57% Indels: 457
DB: 10 Gaps: 1

US-09-920-705-3 (1-901) x US-09-920-705-6 (1-3715)
Qy 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
Db 1832 ATGGCGTTATCTATGAAGCATTCAGAGCCGATGATTCGGTTTCCAGTCAAAACATCTT 1891
Qy 21 ThrValGlySerleuThrProGluThrThrGlnSerAspLeuThrGluLeuPheGlyArg 40
Db 1892 TGGGTGCGTAGCCTTAACGCGGAGACGACAGAGTCAATCTGACCCAGTGTGTTGAGACA 1951
Qy 41 TyTGlyAspIleAspArgIleThrValTySerSerArgGlyPheAlaPheIleTyTyP 60
Db 1952 TAGGGGATATTATAGATCAACGAGTGTATCTTCCAGAGCTTTCGTTTATATATAC 2011
Qy 61 ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
Db 2012 AGACATGTGGAGAGAACAGTCCGACGCCAAAGAGCTCTTCAAGAGCAAAATTTGAATGA 2071
Qy 81 SerGlnIleLysIleGluTyraAlaArgPro----- 90
Db 2072 AGTCAAAATTAAGTCAATACGACGACGCGTTTGTCTTATCTATATCTTGCTTGTTC 2131
Qy 90 ----- 90
Db 2132 TCTAATTGATTGTCTTTTGTCAAGATTATATCTTTTGGGAATTCAATAGTCCAGGT 2191
Qy 90 ----- 90
```

```
Db 2192 TCACAAACTTTGATGATGCTGTTTATGTCACAAAATTTCTTGTAATCTGTTTTTTT 2251
Qy 90 ----- 90
Db 2252 TTCCTATCGTACAAATCAAGTCGAAACCTAGTTTTTTTCTATATATAGTGTAGCTT 2311
Qy 90 ----- 90
Db 2312 AAGGCGAAACCTGATCCGATCGAAACGTCCTTTCGAATTAATCTTGTGTTATATCGAAGT 2371
Qy 90 ----- 90
Db 2372 CCGCGAAAGCCAAACACAGAGAGCTCTGCAAAATTTGATGTAAAGCATATATATCTC 2431
Qy 90 ----- 90
Db 2432 TTAGCAATGAGCTCTGCAAAAGAAATACATCAACACATGTTTACTCTGTTATGCGAA 2491
Qy 90 ----- 90
Db 2492 GAAGTTTAATCCGATGTCGTTTATCTGCAAACTGTATCCGCGTTAGCAGTCTGAT 2551
Qy 90 ----- 90
Db 2552 TCAGTTGCGCTTCAGATGTTAAATCTCACAGCTTGAGATGAATGTAATGGTACTCCA 2611
Qy 90 ----- 90
Db 2612 CTCGAAATTCGCAATGTGTGAAATTTGGAAGCGACCAACTCTATCATCGGCCAAACAGA 2671
Qy 90 ----- 90
Db 2672 TTAAGAAATTTGGAGTTTAAAGTTCAAGTTTCCTGCAAAATCAAAACCGGTGAGAAAT 2731
Qy 90 ----- 90
Db 2732 TTGTCTATGGCGGTACGTAGATATCATATATCATGATCTGCCAAACGAAATTTCTGGCA 2791
Qy 90 ----- 90
Db 2792 CCGATATGTCCTTAACCTCTAGTATGCTGAGACATATCATATCATATGATGAGAGATTTT 2851
Qy 90 ----- 90
Db 2852 GGGCTTGTGTTGGCTCATGCTTCAAGAAACAAATTAAGTGTGTTAAGGCGGCTAAC 2911
Qy 90 ----- 90
Db 2912 TCTACCAATCAGAAACGCTTATTCGAGAAACCATGTTTGTTCCTCAATTCCTATCCCTA 2971
Qy 90 ----- 90
Db 2972 CGTACATCTGGGCTTTCCTCATGTTAGTTCCTTAGAAGTGTGACTTCTTCAACCATTC 3031
Qy 90 ----- 90
Db 3032 TGGGATTGTACAGTTGACGTGACAGACAAATATTCATTGTCAGGACGACACTCAATTT 3091
Qy 90 ----- 90
Db 3092 CAATGCTTGGCGGCTTGAATTAATGGAATGTTTCAAGGATGGTTTATTTGGGAC 3151
Qy 90 ----- 90
Db 3152 TCAGGCATGTATTAAGACAGAAATGTTTTCAGAAAGTATTAATGTCCTTCCGCTTGG 3211
Qy 90 ----- 90
Db 3212 GGGACTATATCTTCTCTACAGCAAAATCTGATGTTTAGGCAAGAGAACTATTAACA 3271
Qy 90 ----- 90
```


Db 1306 GGAACACAGTTG-----GATAGCAGCTTGCAGAGAGTGCATGTATGATCCCA 1353
Qy 867 AsnGlnArgTyrGlnSerThrLeuGlnPheAlaIAsnLeuLeuGlnIleGln 885
Db 1354 CAGAGAGCTGTACAGCAACACATGACGTGGCGCTGTCTTCTTCCACCAATCCAG 1410

RESULT 12
US-10-424-599-40095
; Sequence 40095, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: Ia Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 40095
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT3847_136205C.1
US-10-424-599-40095

Alignment Scores:
Pred. No.: 1.5e-30 Length: 1631
Score: 447.00 Matches: 155
Percent Similarity: 42.80% Conservative: 59
Best Local Similarity: 31.00% Mismatches: 195
Query Match: 9.31% Indels: 91
DB: 17 Gaps: 17

US-09-920-705-3 (1-901) x US-10-424-599-40095 (1-1631)

Qy 432 ArgGlyProProAspSerAspHisIle-----TTrpArgGlyMetIle 445
Db 49 AGATTCAACGCTGTAATCGATAGTTCATCTTAACTGAGTGAATGGAAAGCAACATT 108
Qy 446 AlAlayeglyGlyThrProValCysCysAlaArgCysValProMetGlyGlyIleGlu 465
Db 109 GCTAAAGCGGGAACCCCTGTTGTTGTGACAGCTGCTCCCTGTTGGAAAGTCTAGAT 168
Qy 466 ThrIleuPheProGluValAlaAsnCysSerIlaArgThrAspLeuAsnMetLeuAlaLys 485
Db 169 ATGATGTTACTCTGAGTCTTGATTTGACATGCAAAAACCTGCTTAAGATGCTTTCAAG 228
Qy 486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluAsp 505
Db 229 CATTAACACCAAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 288
Qy 506 PheAlaSerTyrThrGluPheLeuArgTyrIleuSerSerIyAspArgAlaGlyValAla 525
Db 289 ATGGAAGCTTAACATGATTAATGATATCTGAGAGAAAGAACCGCGCTGACGTGCC 348
Qy 526 LysLeuAspArgGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspVal 545
Db 349 AAGTTGATGACAAACACCACTTAATTTCTGTTCTCCATCAGAGTTCTCAGAGAAAGTA 408
Qy 546 LeuGlnArgTyrThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProAla 565
Db 409 CTGAAGGTAAGTCTGAGTGAAGATATCTGTTATCTGAGGTTGAGAACCTCGAT 468
Qy 566 ValProValThrAlaSerTyr-----ArgGlnIleuSerGlnSerAsnProLeuHisTyr 583
Db 469 TTAATATCATGCTCTGAACACATTCAAAAGAAATGACTAAACAAATCTGCTCTAT 528
Qy 584 MetAspGlnAlaArgAspSerProAlaAsnHisSerHis-----SerLeuTyrProPro 601

Db 529 AATGAG-----AATATCTGCTATCTGAAGCTATCATTTCTTCA 567
Qy 602 ---ArgGluAsnTyrIleArgGlyAlaProGlnHisLeuThrAla----- 615
Db 568 GTACGTGTAACCTACATTTCCATCTATTCGAAATGGGTAAATCGAATCAGCAATCTT 627
Qy 616 -----AlaSerLysProSerValSerGluProLeuArgIlePro 628
Db 628 TCATTTCTGGAAATAAATTTGAGACAGCTCCATCAGTTCTGATTCAGCTCGT----- 661
Qy 629 AsnAsnAlaIleProGlnAlaGlyValSerLeuThr----- 640
Db 682 -----GCTGCGCTAGCATGCTCCGAGTTTCATGATGAAGAACGCTAC 723
Qy 641 -----ProGluLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAla 658
Db 724 AACTATCTTACCCAGCAGCGTACTTCTTCAAAATCTCCAGAACTTCTTAATAAGCT 783
Qy 659 AlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGln 678
Db 784 CTACCA-----TTGAGCGCATCAAGTGTGCTGAGAACCTTATGCTGATGAACACCA 837
Qy 679 SerAsnGlyLeuTyrAsnGlyAlaIleProSerGlnAlaIleTrpLysArgGlyProGlnThr 698
Db 838 -----CTATCATTTCCAAAGCGCAGCGCAGAT 864
Qy 699 Val-----HisAspAlaSerAsnGlnSerPheGlnIleTyrGlyAsnGlnTyr 714
Db 865 GTAAAGCGCATTCACATCCCATCCAGTGAATTTCTGATTCATTTATGATGAT----- 918
Qy 715 ThrProAlaGlyGlnLeuProProProProSerArgTyr-----Pro 728
Db 919 -----AGCAAGTGTCTTATCCAGATATGAAGATTAATGATCCTTATCAGTGCCT 969
Qy 729 ProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGln 748
Db 970 GTGGAGGCTCTCTCCAGAACAGCTTGCAACATGACGCAACAGCTTCTTGACGACGAG 1029
Qy 749 SerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyr 768
Db 1030 AGGCATTCGGAGAGCTCTTGCATACATCTGCGTGGCTGAT---CCCAGCAGATTAAAC 1086
Qy 769 SerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyr 788
Db 1087 AGCTTTGGCACAATCCGACACCTCATCCAGATTATCTCA----- 1125
Qy 789 GlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAla 808
Db 1126 TCCGACAACTCATTCAGACACCCCGAAATATGCTACAGAAACATGTGGTAATTC 1185
Qy 809 AsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAla 828
Db 1186 GATCTCTACGATCTGATGGGTCAAAAGTCAAAATGCTACAGTGAAGAACCAACAG 1245
Qy 829 AlaMetGlnProPro-----AlaAspLysAlaAsnLeuGlnProGlnAsnGlnAlaLeu 846
Db 1246 ATCGTAAATGTGCTCAGTGTGTCACAGATAGTTCAAAAGGAAACCAAAAGAGGCGCAAT 1305
Qy 847 ArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLys 866
Db 1306 GGAACACAGTTG-----GATACAGCTTCCAGAGAAATGATCATGTATGATATCC 1353
Qy 866 AsnGlnArgTyrGlnSerThrLeuGlnPheAlaIAsnLeuLeuGlnIleGln 885
Db 1354 ACAGAGAGCTGTACAGCAACATGACGTGGCGCTGTCTTCTTCCACCAATCCAG 1411

RESULT 13
US-10-425-114-31541
; Sequence 31541, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong


```
QY 544 AspValIleuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProPro 563
Db 189 AATGTGTTACAAAGTGTATGTCGACAGCGCTTATGTGTGTATTCACATTCACAA 248
QY 564 ProAlaValProValThrAlaSerTyrArgGlnGlnSerGlnSerAsnProLeuHisTyr 583
Db 249 ATCTGTGCTGTGCTGTGCTGACAGGACAGTTAACTGGACAGAACAGCAACTTAC 308
QY 584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu 603
Db 309 TATGATGAAGGGAAGAACTGTGCTACACAAAGAAAGTATAGTATCATTTCTCTCT 362
QY 604 AsnTyrIleArgGlyAlaProGlnHisLeuThrAlaAlaSerLysProSerValSerGlu 623
Db 363 -----ATGGCAGTGGCCACTTGATGCTGATTAATCGGCGATCTTTGCATGAG 410
QY 624 -----ProLeuArg-----ProLeuArg----- 626
Db 411 GATTCAATGCACGTTGGGGCATATTCTGGAAGGCGCTCGGGTGATGAAGGCCAAGCA 470
QY 627 -----IleProAsnAsnAlaAlaPro-----GlnAlaGlyVal 637
Db 471 GTCCAGCCAGCTCTTGAGGTTTCTGCTAACCAAGCAACTGGGCTGCAAGTGCATCT 530
QY 638 SerLeuThrProGlnLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnPro 657
Db 531 TCACTCAAGCCTGACATATGCTACTTAGCAAAACTCTGCTAGTGTG----- 581
QY 658 AlaAlaProGlnSerHisGlnProMetSerGly 668
Db 582 -----CAATCATCAACAAGTGTAGTGT 605
```

Search completed: April 8, 2005, 16:35:33
Job time : 1297 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 08:31:28 ; Search time 6977 Seconds

(without alignments)
4915.566 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801
Sequence: 1 MALSMKPRFRDSDGFSNNL.....LQIQKQKQSSGSGTPAGQGP 901

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool.h/US09920705/runatc_07042005_101813_20668/app_query.fasta_1.1095
-DB=EST -OPMT=fastcap -SUFFIX=rcst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pcc -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09920705_@CGN_1_1_4087_@runatc_07042005_101813_20668 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOC=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4263.5	88.3	3082	3	CNS0A83E
2	328.5	19.3	767	8	BH552322
3	808.5	16.8	866	9	CL511524
4	751	15.6	724	8	BH552335
5	747	15.6	812	5	BX840244
6	716.5	14.9	2706	3	CNS0A2CE
7	695.5	14.5	602	4	BG543956
8	693	14.4	701	2	BE037904
9	607	12.6	635	1	AU237346

10	561	11.7	416	8	BH851246
11	537	11.2	582	1	AV829561
12	535	11.1	817	7	CO095058
13	532.5	11.1	800	6	CA766318
14	532.5	11.1	1105	7	CK208950
15	511.5	10.7	582	7	CF322873
16	510	10.6	685	5	BH831648
17	507	10.6	412	8	BZ763896
18	500	10.4	281	6	CB260919
19	493	10.3	655	2	AM696401
20	486	10.1	630	4	BG447636
21	470	9.8	295	9	AL938124
22	461	9.6	362	8	BZ288571
23	454.5	9.5	624	5	CA032838
24	452	9.4	561	2	BF266387
25	451	9.4	617	4	BG649121
26	450.5	9.4	683	3	CK851161
27	450.5	9.4	855	6	CB656009
28	448	9.3	662	2	BF641487
29	441	9.2	421	5	BP647781
30	430.5	9.0	441	6	CD481691
31	423	8.8	647	5	BQ696544
32	418	8.7	542	7	CO910873
33	409	8.5	546	7	CN893760
34	406.5	8.5	754	6	CA205899
35	406.5	8.5	793	9	CC634984
36	403	8.4	493	2	BF639287
37	400.5	8.3	814	6	CB633235
38	399	8.3	278	9	AL938125
39	398	8.3	685	4	B1954416
40	396	8.2	950	9	CL241131
41	385	8.0	831	7	CNE21077
42	384	8.0	477	7	CF759151
43	379	7.9	854	9	CG960265
44	377	7.9	381	5	BQ697581
45	376	7.8	666	7	CV092674

ALIGNMENTS

RESULT 1
CNS0A83E
LOCUS
DEFINITION
CNS0A83E
Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSUTP660202 of Hormone Treated Callus of etratin col-0 of Arabidopsis thaliana (thale cress).

ACCESSION
BX820695
VERSION
BX820695.1
KEYWORDS
HTC; GSUT cDNA.

SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)

REFERENCE
AUTHORS
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

TITLE
1 (bases 1 to 3082)
Castelli V., Aury J.M., Jallion O., Wincker P., Clapet C., Menard M., Craud C., Quetier F., Scarpelli C., Schachter V., Temple G., Caboche M., Weissenbach J. and Salanoubat M.

Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 3082)
Genoscope.
Direct Submission
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

COMMENT
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Craud C.,

Schachter V., Weissenbach J., Salanoubat M.

URCV INRA : Clépet C., Caboche M.

Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap.

http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full

length

<http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis>.

FEATURES

SOURCE

location/Qualifiers
1..3082
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
/db_xref="taxon:3702"
/clone="GSLTPGH60ZD02"
/issue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT_6"
complement(1..3082)
/gene="At2g43410"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	3082
Score:	4263.50	Matches:	840
Percent Similarity:	92.74%	Conservative:	3
Best Local Similarity:	92.41%	Mismatches:	15
Query Match:	88.80%	Indels:	54
DB:	3	Gaps:	1

US-09-920-705-3 (1-901) x CMS0A83E (1-3082)

```

QY 1 MetAlaLeuSerMetLysPro--PheArgAlaAspSerGlyPheGlnSerAsn 20
Db 234 ATGGCGTTATCTATGAAGCATGTGCAAGCCGATGATTCGGTTCCAGTCAAAACATC 233
QY 20 eutrpValGlySerLeuThrProGluThrGluSerAspLeuThrGluLeuPheGlyA 40
Db 294 TTGGGTCGGTAGCTTAACCGCCGAGACGACAGATCAGATCTGACCGAGTTTGGAA 353
QY 40 rgtYrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyr 60
Db 354 GATACGCCGATATTTGATGAAATCAAGGTATTTCTTCAAGGCTTTGCGTTATTAATCT 413
QY 60 YrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsn 80
Db 414 ACAGACATGTGGAAGAGCAGTCGACGCAAGAGGCTTTCAAGGCAAAATTTGAAG 473
QY 80 LysArgGlnIleLysIleGluTyrAlaArgProAlaLysProGlySerLeuTyrValG 100
Db 474 GAAGTCAAAATTAAGATCGAATACGACGACGCAAAACCTTGTAAGATCTATGGGTG 533
QY 100 LysGlyIleGlyProAsnValSerLysAspAspLeuGluGluLysPheSerGlyLe 120
Db 534 GTGAATCGGCCCTTAATGTCTCAAGATGACCTGAGAGAAAGTTCAACAAGTTGGGA 593
QY 120 YrIleGluAspPheLeuArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGlu 140
Db 594 AAATTCAGAGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTCATGTATTTATGAGA 653
QY 140 eLAspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeu 160
Db 654 TGAATATGCTTTACAGGCTTAAGACATGATGAAAGCTATGGGTGATGCTTTTTC 713
QY 160 rgtYrAlaPheLeuArgSerGlnAlaProLysLysGluGlnTyrAlaGlySerTyrAsp 180
Db 714 GTGTGATTTTTCCTCGGTCAAGCGCCAAAATAAACAATGCGCTGCTCTTAAGATA 773
QY 180 snAspGsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysG 200
Db 774 ACAGAATATGCAATATGAATCATAAACCCACAGATCTCATCATATGAAAGCTTTAAAG 833
QY 200 LysAspValGlnProSerLysValLeuTyrIleGlyPheProProThrAlaThrGlnCys 220

```

```

Db 834 GAAATGTCACGCCAAGTAAGTTCTGTGATTTGGTTCCCTCTTACTGCTACAAATGCA 893
QY 220 snAspGlnGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLys 240
Db 894 ATATATGAGCAAAATTTGCAACATGCAATGCAATATCTTTGTGTAGATCGAAGGTTAAA 953
QY 240 eTyrTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnC 260
Db 954 GTTACCCATCAAGAAATTTTGACATTTGTGAGATTAGAGCCGCGGAAGCTCCGCAT 1013
QY 260 YrLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSer 280
Db 1014 GCAGGAAGGCTTACAGGGAGAGTTTATCAATTAATCTTGAATCAAAATTAATGACTCA 1073
QY 280 snAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSer 300
Db 1074 ACAGATAGTTGCCCTCCTGAGCAAGACGATACATGATTTTACTCT----- 1117
QY 300 rgtThrAspMetPheAsnAsnProSerCysValSerSerProHisSerThrGlyIleP 320
Db 1117 ----- 1117
QY 320 rgtLysSerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlyTyr 340
Db 1118 -----GAATACA 1124
QY 340 snAspValAlaGlyLysGluProAsnTyrPArgArgProSerAlaAsnGlyThrGlyIle 360
Db 1125 ATGACGTTGTTGTGAAGAGCCAACTGAGAGGCCATCTGCAATGAACTGGAATAC 1184
QY 360 euProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgAspPro 380
Db 1185 TCCCATCTCAACAGACAGCTGGAATCTCCATCTCTGCAACAGAGTCAAGGCCCTTA 1244
QY 380 eLArgSerAsnProAspSerTyrGluGlyTyrAspProAlaGlnLeuValArgGluSerL 400
Db 1245 TGAGGTCAAAACCCGATTTCTTGGAAGATATGATCTGCTCAAGTTGTCAGAGAACTA 1304
QY 400 YrArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluAs 420
Db 1305 NACGAACCAAGAAATGATGATCAGTGAACGGATTTTACCCNAAGGCTGCTCATGAGAGT 1364
QY 420 ePheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHis 440
Db 1365 CATTTGTCGAGGTTCAATGCTGCTGATACCTATCCGAGGCCCTCTGATTTCTGATCA 1424
QY 440 lETPrArgIlyMetIleAla-LysGlyGlyThrProValCysCysAlaArgCysValPro 459
Db 1425 TATGAGAGGAATGATTCCAAAGGGTGAATCTCCGCTGTGTGCTCGATGTGTACT 1484
QY 460 MetGlyLys-GlyIleGluThrLysLeu-ProGluValAlaAsnCysSerAlaArgThr 479
Db 1485 ATGGGAAGGGGAGATGNAATCAACTGCTCGTAGGTGTCAATTTGGTCCGACAGAACTG 1544
QY 479 sPLeuAsnMetLeuAlaLysHisSerTyrAlaValAla-LIleGlyCysGlnIleValPhePhe 498
Db 1545 ATTTGAATATGCTGCTTAAACATTACCGCGTTGCAATGGAGATGAGATGTTTTC 1604
QY 499 ValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSer 518
Db 1605 GTACCAAGCAGAGAAAGATTTTGCGCTTACACTGA-TTTCTTCGATCTTTAGGCTCA 1663
QY 519 LysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProPro 538
Db 1664 AAAGATCGGGG-GGTGTTGCCAAATTTAGATGATGATCAACTTATTTCTGGGCTTCA 1722
QY 539 SerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValVal 558
Db 1723 TCAATTTCTTAACATGATGATCTCAAGTGAACCGTCAAGAGCGCATATATGGGT-GTT 1781
QY 559 LeuLysLeuProProProAlaValProValThrAlaSerTyrArgGlnGlnSerGlnSer 578
Db 1782 CTCAAGTTACCCCGCAGCGGTCTGTTCAGATCATATACAGACAAAGATCTCACTCC 1841

```

```

Qy 579 AsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeu 598
    |||||
Db 1842 AATCTCTGATATATGATATCAAGCCCGGATTCACCTGCCAATGGCTACAGTTTA 1901
Qy 599 TyrProBraArgGluuAenTyrIleArgIValaProGluHisLeuThraAlaAserLys 618
    |||||
Db 1902 TATCTCTCTAGGGAATTAATCATAGGGGGTGCACAAACATTGAGACGGCTTCAAA 1961
Qy 619 ProSerValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSer 638
    |||||
Db 1962 CCATCTGTAGCGAGCTCTCAGAAATCTTAATATCAGCGCTCAAGCTGGGGTTAGT 2021
Qy 639 LeuThraProGluLeuAlaThraLeuAlaSerIleLeuProAlaThraSerGlnProAla 658
    |||||
Db 2022 TTAATCTCGAGGCTTTTAGCCACTTGCCACTATATCTCCCTGCACCTTCTCAACCTGCT 2081
Qy 659 AlaProGluSerHisGlnProMetSerGlyProSerThraValaSerThraAlaHisGln 678
    |||||
Db 2082 GCCCTTAGAGTGCACCAACTATGTCAAGACCTTCAACAGTTGTTTCCACAGCATCAG 2141
Qy 679 SerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrrpLysArgIleProGlnThr 698
    |||||
Db 2142 TCCATATGACTGTACAAATGGAAGACACCGCTCAAGCTTGGAAAGAGGTCCCAACA 2201
Qy 699 ValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThraProAlaGly 718
    |||||
Db 2202 GTTCATATATGCGTAAATCAAGTCAATTCACAAATACGAAATCGTACACTCCAGCTGGG 2261
Qy 719 GlnLeuProProProProSerArgTyrProProAlaSerAsnProAsnTyrThraSer 738
    |||||
Db 2262 CAACTACTCTCTCTCTGCGGTACCTCCAGCTTCAAAACACCCCACTACACTAGT 2321
Qy 739 GlyMetValHisGlyAsnMetGlnTyrGlnSerGlnSerValaAsnMetProGlnLeuSer 758
    |||||
Db 2322 GGAATGCTCATGCGCAACATGCAATACAGAGCCAACTGTTAACTGCTCAGCTGTCT 2381
Qy 759 ProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThraGlnGlySerSerAsnHis 778
    |||||
Db 2382 CCGTATCCAAATATAGCTCATAAATATTTCCATGTAACCTCGGGTGTGCTCAATCAT 2441
Qy 779 ProValSerGlnProMetValaGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsn 798
    |||||
Db 2442 CCGTCTTCTCAGCCCATGCTCCAGCATACCAACCAAGGCTCATGCTCCAAACCAAAAC 2501
Qy 799 TyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThraAsnGlnAla 818
    |||||
Db 2502 TATGTTCCAAATTCAGATTATCAGCAAGCTAATTTTCATGGCGTAAACAATCAGGCA 2561
Qy 819 GlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysValaAsn 838
    |||||
Db 2562 CAGAACTTAAACCTTCCCAATTCAGCTGCCATGCAACCAACAGACAGATTAAGGCAAT 2621
Qy 839 LeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGly 858
    |||||
Db 2622 TTATAGCCAAACCAATCATCATGAGTTGAGCTTATGATCTTGGGAGATGGTCATGGG 2681
Qy 859 -ThraAspGlyGluValaAspLysAsnGlnArgTyrGlnSerThraLeuGlnPheAla 878
    |||||
Db 2682 TAAACAACAGATGGGAGGTGCAATAGATCAGAGATCAGAGTCAACACTCAATTTTCAG 2741
Qy 878 IAsnLeuLeuLeuGlnIleGlnGlnIleArgIleGlnGlnGlnSerSerGlyThraProAla 898
    |||||
Db 2742 CAAACCTTCTTCTCAGATACAGCAAAACAGCAAGAGTCTTTAGGTAATCCGGCTG 2801
Qy 898 LysGlnGlyPro 901
    |||||
Db 2802 TACAGGGGCT 2812

```

```

RESULT 2
BH552322 767 bp DNA linear GSS 14-DEC-2001
LOCUS BH552322
DEFINITION BOGHQ3TF BOGH Brassica oleracea genomic clone BOGHQ93, genomic

```

```

ACCESSION BH552322
VERSION BH552322.1 GI:17804102
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE 1 (bases 1 to 767)
AUTHORS Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOGHQ3TR
Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
    source
        1..767
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOGHQ93"
            /note="Vector: pHOSt, Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOSt using BstXI linkers"

```

ALIGNMENT

```

Alignment Scores:
Pred. No.: 3,096-64 Length: 767
Score: 928.50 Matches: 198
Percent Similarity: 74.64% Conservative: 8
Best Local Similarity: 71.74% Mismatches: 37
Query Match: 19,34% Indels: 33
DB: 8 Gaps: 9

```

US-09-920-705-3 (1-901) x BH552322 (1-767)

```

Qy 499 ValProAspArgGluGlnuAspPheAlaSerTyrThraGluPheLeuArgTyrLeuSerSer 518
    |||||
Db 751 CTACAGACAGCGAAGAAATTCGCTTCTTACACTGAGTTTCTCGCTACCTGACTCG 692
Qy 519 LysAspArgAlaGlyValaLysLeuAspAspGlyThraThraPheLeuValaProPro 538
    |||||
Db 691 AAAGATCGAGCTGGGGGATGCAATTAAGAGATGAACATTAATTTCTGCTGCTCCA 632
Qy 539 SerAspPheLeuThraAspValaLeuGlnValaThraArgGlnGluArgLeuTyrGlyVala 558
    |||||
Db 631 TCAGATTTCTTAATCTAATGATCTCAAGTACCGGCTCGAGGCTCTTAAGCGCTTGT 572
Qy 559 LeuLysLeuProProProAlaValaProValaThraAlaSerTyrArgGlnGlnSerGlnSer 578
    |||||
Db 571 CTCAGACTGCCCCACACAGCGCTCCGTTGACGATCGTACAGACAGAGTCTCAGTCT 512
Qy 579 AsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeu 598
    |||||
Db 511 AACCTCTGTCTTAATATGATTAACCCCGATTCCTGCCAATACCGGTACACCTTT 452
Qy 599 TyrProBraArgGluuAenTyrIleArgIValaProGluHisLeuThraAlaAserLys 618
    |||||
Db 451 TATCTCTCAGGGGT-----GCTGCTGCACCAAGACAG-----TCAACA 413
Qy 619 ProSerValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValaSer 638
    |||||
Db 412 CCATCGGTTAGCAGGCTCTCAGGTTACTTAACAATGACAGCGTCAAGCTGGGGTTAGT 353

```


TITLE Whole genome shotgun sequencing of *Brassica oleracea*
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGH093TF
Contact: Chris Town

712R
712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@icgr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
Source

Location/Qualifiers
1..724
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGH093"
/clone_lib="BOGH"
/note="Vector: PHOS1; Site 1: BexXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BexXI linkers"

ORIGIN

Alignment Scores:
Pred. No.: 5 22e-50 Length: 724
Score: 751.00 Matches: 150
Percent Similarity: 70.29% Conservative: 18
Best Local Similarity: 62.76% Mismatches: 24
Query Match: 15.64% Indels: 47
DB: 8 Gaps: 4

US-09-920-705-3 (1-901) x BH552335 (1-724)

QY 173 GlnTPrAlaGlySerTyrAspAsnArgAsnGlyAsn---MetAsnHisLysProGln--- 190
DB 55 AAATGGCTGGCTCTTAAGATTAAGAAATGGCAATGGATGATCAAAACCTCAGATT 114
QY 190 ----- 190
DB 115 CGTTTAAATATTCAAAAAAGACTTGTATTGTTCTGCTTATATGTATGATGATT 174
QY 191 -----TyrProHisSerTyrGluAspPheLysGlyAs 201
DB 175 AATTAGCTTTTGAAGTTTGTGTATGATCTCACTCACTCAATGAAGAAGCCAGAGAGA 234
QY 201 pValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAs 221
DB 235 TGAACCAAGCAAGTAAAGTTCTGTGATTTGGGTACCTCTCTCTTCTGAG-----ATAGA 288
QY 221 pGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyr 241
DB 289 CGAGCAATGCTCAACAACCAATGATATCTTCTGAGAGATCGAGAGAGAAAGTAA 348
QY 241 rProSerAraAspPheAlaLeuValGluPheArgSerAlaGluGlnAlaArgGlnCysL 261
DB 349 CCGGTGAGGAGCATTTTCACTGTGAGATTGAGAGGCTGAGCAAGCTGCCAAGCCAA 408
QY 261 sGluGlyLeuGlnGlyArgLeuPheAsnAspProArgIleLysIleMetTyrSerAsnAs 281
DB 409 GGAAGGCTCAAGAGAGGAGGCTATTCAAGGATCCAGATCACTATCTATCTGAGAGA 468
QY 281 pGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgTh 301
DB 469 TGAATATCCTCTCTGAGCAAGATGATTAAGTTTATTTCTGCGCTGAACGGTCAAGGCC 528
QY 301 rAspMetPheAsnAspAspProSerCysValSerSerProHisSerThrGlyIleProG 321
DB 529 AGGTATATTCATCACTATCTCT----- 550
QY 321 ySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValaGluTyrAsnAs 341
DB 551 -TCAATTAGCCATTCAGAGGTAGCATGACCGTTCTTATATATGTTTCAAGATTACATGA 609

QY 341 pValAlaGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuPr 361
DB 610 CGTTGTGTATGAGCAAACTGAGAGAGCGCTCTCCAAATGAGACTGGGATATCTCC 669
QY 361 sSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgPro 379
DB 670 ATCTCCAGCAGACATGGATTCCTCTCTCTCTGCAAGAGATAGAAACCT 724

RESULT 5
BX840244
LOCUS BX840244
DEFINITION BX840244 Arabidopsis thaliana Hormone Treated Callus Col-0
ACCESSION Arabidopsis thaliana CDNA clone GSUTPH87ZF04 SRIM, mRNA sequence.
VERSION BX840244.1 GI:42534327
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS
Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C., Menard,M., Cnaud,C., Quetier,F., Scarpelli,C., Schacher,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
Unpublished (2004)

JOURNAL

Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr

TITLE

The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aury J.M., Jallion O., Wincker P., Menard M., Cnaud C.,
Schacher V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences).
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/EST
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

COMMENT

Location/Qualifiers
1..812
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="GSUTPH87ZF04"
/tissue_type="Hormone Treated Callus"
/clone_lib="Arabidopsis thaliana Hormone Treated Callus
Col-0"

FEATURES

Source
1..812
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="GSUTPH87ZF04"
/tissue_type="Hormone Treated Callus"
/clone_lib="Arabidopsis thaliana Hormone Treated Callus
Col-0"

ORIGIN

Alignment Scores:
Pred. No.: 1.3e-49 Length: 812
Score: 747.00 Matches: 148
Percent Similarity: 86.84% Conservative: 17
Best Local Similarity: 77.89% Mismatches: 23
Query Match: 15.56% Indels: 2
DB: 5 Gaps: 0

US-09-920-705-3 (1-901) x BX840244 (1-812)

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
DB 246 ATGGCTTATCTATGAAGCATTCAGAGCGGATTCGGTTCAGTCAAAACAATCTT 305
QY 21 TrpValGlySerLeuThrProGluThrThrGlnLysSerAspLeuThrGluLeuPheGlyArg 40

Db	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Db	306	TGGGTGGTGGCTTAACGCCGGAGACACAGAGCTGAGATCTGACCGAGTTGTTGGCAGA										
Qy	41	TyrGlyAspIleAspArgIleThrValTyrSerSerArgIlyPheAlaPheIleTyrTyr										
Db	366	TAAAGCCAAATTGATTAGACTCCGGGTGATATTTCTTCACGAGGCTTTGGCTTATCTCTAC										
Qy	61	ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlnIlyAlaAsnLeuAsnGly										
Db	426	AGCATCTGGACGACAGCATGTCGACGCCCAAGAGCTCTTCAAGAGCAAAATTTGAATGGC										
Qy	81	SerGlnIleIysIleGluTyrAlaArgProAlaIlyProCysIysSerLeuTyrValGly										
Db	486	AGTCAAAATTCAGATTGATACGATACGCAAGACCGGCAAGCTTGTCAGATCATGGTGGGT										
Qy	101	GlyTlIGlYProAsnValSerIlyAspAspLeuGlnGluGluPheSerIlyPheGlyIly										
Db	546	GGCATGGCAGCTATGATGATCCAAAGATGAACAGAAAGCAAAAGTTGACGAATTTGGAAA										
Qy	121	IleGluAspPheArgPheLeuArgIlyGluTyrGlyThrAlaPheIleAspTyrTyrGlu										
Db	606	GTCAGAGATTTTACGGGTTCTTCAGAAAACGAAGACAGATTTCTTGATTATTCATGAGA										
Qy	140	LeuAspAspAlaLeuGlnAlaIlySerSerMetAsnGlyIlyPheGlyIlySerPheLeu										
Db	666	TGATATATGATTTTACAGATTAAGAGATGATCTGGGAAGCCACCGGTGGCTGCTTTTTC										
Qy	160	ArgValAspPheLeuArgSerGlnAlaProIlyLeuGlnGlnTyrAlaGlySerTyrAsp										
Db	726	GTGTTAAATTTCTCCGGACACCAAGACCAAAAGGAGAAAAATGGCTGACTTACGATA										
Qy	180	SerArgAsnGlyAsnMetAsnHisIly										
Db	786	ACGAAATGCAAAATTAATTAATTA										
RESULT 6	CNS0A2CE	2706 bp mRNA 1linear HTC 04-FEB-2004										
LOCUS	Arabiidopsis thaliana Full-length cDNA Complete sequence from clone											
DEFINITION	GSUFB492D02 of Flowers and buds of strain col-0 of Arabidopsis											
ACCESSION	Chailana (thale cress).											
VERSION	BX842306											
KEYWORDS	BX842306.1 GI:42406938											
SOURCE	HTC; GSUFT cDNA.											
ORGANISM	Arabiidopsis thaliana (thale cress)											
REFERENCE	Arabiidopsis thaliana											
AUTHORS	Eukaryota; Vitridiplantae; Streptophyta; Embryophyta; Tracheophyta;											
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;											
JOURNAL	rosids; eustosids II; Brassicales; Brassicaceae; Arabidops.											
AUTHORS	1 (bases 1 to 2706)											
TITLE	Caselli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,											
JOURNAL	Menard,M., Cruaud,C., Queller,F., Scarpelli,C., Schachter,V.,											
AUTHORS	Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.											
TITLE	Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences											
JOURNAL	A Combined Approach to Evaluate and Improve Arabidopsis Genome											
AUTHORS	Annotation											
TITLE	Unpublished											
JOURNAL	2 (bases 1 to 2706)											
AUTHORS	Genoscope.											
TITLE	Direct Submission											
JOURNAL	Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :											
AUTHORS	BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr											
TITLE	- Web : www.genoscope.cns.fr)											
JOURNAL	The sequences are based on single pass reads.											
AUTHORS	Life Technologies (a division of Invitrogen) members carried out											
TITLE	full-length libraries construction : Temple G.											
JOURNAL	Genoscope members carried out sequencing and annotation : Caselli											
AUTHORS	V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,											
TITLE	Schachter V., Weissenbach J., Salanoubat M.</											

[illegible]

QY 253 ralaaglualalarglncysylsgluleglnllyargleuphenaenp- 273
 Db 636 tctgattgcagcttggatgagcgaagaaagcttgcaggaagatttggcaattcttcg 695
 QY 273 gilelyllemetyserasnaenp- 282
 Db 696 agtgcattatttggtttgcaggaagcttctctcattgcagcagtggaagggtccgtc 755
 QY 283 -----leuproglu- -----Aspaspthserpethyserglyme 296
 Db 756 gggtagatcactctcacccttaccagatctgttgatgcactggagatcttcagaaagttta 815
 QY 296 tlybargserargthr- -----Aspmetpheaenapserpsercysvalserperphi 315
 Db 816 tcttcagatgaaactatgaaagcagcagaaattccag- -----gtgaggaagaccaca 872
 QY 315 ----- 315
 Db 873 ttacatagaagatgagattttagaagattctgaagttacattttttaaagaaagaga 932
 QY 316 -----Serthrglyleproglisermetargproleuargglythraenglargse 333
 Db 933 ctcaagtagcagatgcagctcctgcattatggaagctc- -----agatctacgatatgatttcc 989
 QY 333 rtyrargnglyalaglutyraenap- valvalgllylsgluproantthpargargpros 353
 Db 990 tcaggaatgacgaaatgacatgagtagctcgggga- -----tggtacttctcatt 1040
 QY 353 exala-----Asnglythrglylleuproserprothrglyproglylle 369
 Db 1041 ccttgatattctcagatctcagacagagagtt- -----cagatgataaagaccatggagattc 1097
 QY 369 eufroserproalagllytharg- -----Argprometargserasnproaspser 387
 Db 1098 acccgaaagatgactactactacacagaaatcagaagattgaaagacagatcctcgagcc 1157
 QY 387 rpgluclytyrarp- -----Glnleu-----Valarggluserlyahargthra 404
 Db 1158 tgagagccagctc-ccagggcagatcagcttctgtgtatgagcaagaaagacgcccctttt 1216
 QY 404 rgaerpglyservalaerpglyphethprometglyvalaerpglyargserphedglyarg 424
 Db 1217 caagggcattcgtcgtat- -----ttttcccaaaagatgatttgacgggaactatgagcgtc 1273
 QY 424 ----- 424
 Db 1274 gacagctaagatcaacaaacagttgagacgctttaaactggtatagaagatggg 1333
 QY 425 -----SerV 426
 Db 1334 acaaagtagctttagcgaacacacatgaaagattatgagagttatcttgcacatcag 1393
 QY 426 alalalalargproleargglyproproasp- -----Seraspheist 440
 Db 1394 ttgtccctgaaagaaagagagacacccctgaactgaatcagacatcctaaagacttga 1453
 QY 440 letparaglymetilealalyeglygllythraprovalcysvalaargcysvalaerom 460
 Db 1454 actggaaagagactatgcttaagagagacacatccatttgcagcaaaatgcttctcgt 1513
 QY 460 etcglylsglyllegluthlyleuprogluvalaenapcysvalaerargthra 480
 Db 1514 tggcgaagatgagatgagatgagctgctgctgatttgcagatggcagaaagactggt 1573
 QY 480 euaenmetleualalyehietryalavalalaleglycysgluilevalphephevalp 500
 Db 1574 tagacatgctggctgaagcatttactacaaatcatatgaagcagtggtgttttcttcttc 1633
 QY 500 rohsparaglygluasphealaserlythrglupheleuargtyrleuaserlysa 520
 Db 1634 ctggaagatgagatgagatgagctgttttattgattgagattttagcatttctggaagagaaac 1693

QY 520 sparaglaaglyvalalalyeaspaapcglythrrhthreuphelevalproprosera 540
 Db 1694 aacggcagagctgttttgaattgagatgacaaacagttgttttctgggtcccttccttcg 1753
 QY 540 spheleuthraerpalaleuglnvalthrargnglnluthrgleutyrglyvalaleul 560
 Db 1754 atttcccgagaaagacttaaaagttcctgggaactaaagcattcctggagatttctac 1813
 QY 560 yaleuproproalavalprovalthralsertyrargnglnuserglnserasn 580
 Db 1814 gttta----- 1818
 QY 580 roleuethiermetlaerpalalargaspaerproalasnaleserhiserleutypr 600
 Db 1819 -----gaatgctggtcttgcctccgggtggc 1846
 QY 600 roproargluanttyrleargglyalaprogluethierleuthralaalaserlypros 620
 Db 1847 ctgttcacagcaagagtagaagaaagatgacagacttgcttaattttagtggaacat 1906
 QY 620 ervalsergluproleuarglyleproasnbnalalaproglinalaglyalserleut 640
 Db 1907 catattcagaa----- 1917
 QY 640 hrprogluueuleualathrleualaserlleuproalathrserglnproalalarp 660
 Db 1918 -----gctagtgagacggttctcga- -----tctgaaacccagatattc 1956
 QY 660 rogluserhnglnprometserglyproserthralvalserthralahleuglnsera 680
 Db 1957 cg-----gctcaaacacttcttgcgctagtgagggcgagaca 1995
 QY 680 englyleutyraenglylualaproserglnalathrlybargglyproglinthvalh 700
 Db 1996 at-----caaaagtcacatgagatgattccatag- -----gtcg 2025
 QY 700 isapalaseranglnserphenglnlthyrglyasnnglnthyrrproalaglygln 720
 Db 2026 agdacaagacatgactgagtagctacccggtattctgaa----- 2063
 QY 720 eufroproproserargtyrproproalaserasnbnproanttyrthr- ----- 737
 Db 2064 -----tcgacgtgcccctcgtgatatcaaaaccccaagatgaccccttca 2109
 QY 738 -----Serglymetvalhnglyasnmetglnthyrglnserg 750
 Db 2110 tagatcatctgttcaaaaacacagcgggttcttctggaaacagaa----- 2159
 QY 750 inservalasmetproglinleuserproleuproasnmetprohlsasnbnthyserm 770
 Db 2160 -----AATGCAATCTAAGTC 2175
 QY 770 ettyrthrglnlyseraserasnleprovalserglnprometvalglnlthyrglnp 790
 Db 2176 gttat-----catgatacgaagaaactccagtcacagaggggttttcagc 2217
 QY 790 roglu-----Alasermetproasnnglnanttyrglyprolepros 804
 Db 2218 ctgaacagctaacacatttagctcctcgttaccagacag-----C 2259
 QY 804 ettyrctglnlmalasnphenieglyvalthrthraenglnalaglnaen-----L 821
 Db 2260 aacaaacagaaacagaa-----acgcgaatgacgcttgagaaatTTTTTTTTT 2307
 QY 821 euaenproserglnphegln-----Alalawetglnproproalalarpvalaasn 839
 Db 2308 tagagcgaagttgttttgcagcatttgcgaacatgagacagacgcaagcattcctcaacg 2367
 QY 839 euglnuproglinaenglnalaleuargleuaglnprometileserglyaerglynglyt 859
 Db 2368 taactccacaaatcaaaatgtacagattaca-----agtaacacacagcacaacac 2418
 QY 859 hrthraerpglygluvalaerlyasnglnargtyrlnserthrrleuglnphealalaa 879


```

Db      2419 AAGAAAGAACAGAGCCCAATCCACAGAGCCTCTTCAAGCAACATTACAGTACACACTG 2478
Qy      879  snleuleuLeuGlnTleGln 886
Db      2479 GACTTCTCCAGCAGATTCAACAA 2501
RESULT 7
BG543956          602 bp      mRNA      linear      EST 01-MAY-2002
LOCUS      BG543956
DEFINITION  E1706 Chinese cabbage etiolated seedling library Brassica rapa
ACCESSION  BG543956
VERSION    BG543956
KEYWORDS
SOURCE
ORGANISM
Brassica rapa subsp. pekinensis
Brassica rapa subsp. pekinensis
Brassicaceae; Brassicaceae; Brassica.
REFERENCE
AUTHORS  Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H.,
          Bahk,J.D., Lee,S.Y., Cho,M.J. and Lim,C.O.
          Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
          Unpublished (2001)
          Contact: Lim, C.O.
          Plant Molecular Biology & Biotechnology Research Centre
          Geongsang National University
          #900 Gajwa-dong, Jinju 660-701, Korea
          Tel: 82 55 751 6255
          Fax: 82 55 759 9363
          Email: colim@nongae.gsnu.ac.kr
          Seq primer: T7.
FEATURES
source
1..602
location/Qualifiers
  /organism="Brassica rapa subsp. pekinensis"
  /mol_type="mRNA"
  /cultivar="Jangwon"
  /sub_species="pekinensis"
  /db_xref="taxon:51351"
  /clone="E1706"
  /tissue_type="Etiolated seedling"
  /lab_host="Xu-1 Blue"
  /clone_1ib="Chinese cabbage etiolated seedling library"
  /note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
ORIGIN
Alignment Scores:
Pred. No.:      1.15e-45      Length:      602
Score:          695.50      Matches:      138
Percent Similarity: 84.02%      Conservative: 25
Best Local Similarity: 71.13%      Mismatches:  26
Query Match:    14.49%      Indels:       6
DB:             4           Gaps:         2
US-09-920-705-3 (1-901) x BG543956 (1-602)
Qy      23  GlySerleuThrProGluThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly 42
Db      13  GGCAGCTGACCCCGGCTACTACGAATCGATCTGCCGATCTGTTCCGAAGGTTCCGC 72
Qy      43  AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
Db      73  GATATGACAGGCTCCACCGCTAATCTTCCGTTGCTTCGCGTTATCTACCTCAGGCGC 132
Qy      63  ValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGln 82
Db      133  GTTGAAGAACCCGTACCGCGGAGAGGCTCTTCAGAGGTCGAATCAACGCGAGTCCG 192
Qy      83  IleValIleGluTyrAlaArgProAlaValProCysValSerSerLeuTyrValGlyGly 102
Db      193  ATTAAATGCAATTCGACGACCGGCAAAACCTGTGAAGAGCTTGGGTTGGTGAATC 252

```

```

Qy      103  GlyProAsnValSerIysAspAspLeuGluGluGluPheSerIysPheGlyIysIleGlu 122
Db      253  AGCTCAGATGTCGCCGAAGATGATCTCGAGGAGGATTCAGAGATTTGGGGAATCGAA 312
Qy      123  AspPheArgPheLeuArgGluArgIysThrAlaPheIleAspTyrTyrGluLeuAspArg 142
Db      313  AGTTTCAGGTTTCTTAAAGATCGCAAGACGCTCTTTATTGACTATTTTGAAGTGAAGAT 372
Qy      143  AlaLeuGlnAlaLysSerMetAsnGlyIysProMetGlyGlySerPheLeuArgValAsp 162
Db      373  GCTGTACAGCAAAAGACCATGAACGGGAAGCCAGTACGATGAGTTTTCGAGTTGAT 432
Qy      163  PheLeuArgSerGlnAlaLeuProLysIysGluGlnTTPAlaGlySerTyrAspAsnArgAsn 182
Db      433  TTTCTTGCTCAACAGGCGCAAGAAAGAACCAACGCGCTTGCTCCGAAGATGCGAGGAT 492
Qy      183  GlyAsnMetAsnHisLys---ProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db      493  GGCAGTTTGAAGTCAGACAGCAGATATCTACTCA-----CTTAGAGAT 540
Qy      202  ValGlnProSerLysValLeuTTPleGlyPheProPormr 215
Db      541  GGCCACAA-AGCAAGTTCTGTGATGGTGGTACCTCCTTCT 581
RESULT 8
BE037904          701 bp      mRNA      linear      EST 07-JUN-2000
LOCUS      BE037904
DEFINITION  A06B09 AA Arabidopsis thaliana cDNA 5', mRNA sequence.
ACCESSION  BE037904
VERSION    BE037904.1 GI:8332920
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (chale cress)
Arabidopsis thaliana
Brassicaceae; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS  Bohner,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M.,
          Ferreira,H., Kawasaki,S., McColough,A., Michalowski,C.B.,
          Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
          Functional Genomics of Plant Stress Tolerance
          Unpublished (2000)
          Contact: Michalowski,C.B.
          University of Arizona
          Bio Sciences West room 513, Tucson, AZ 85721, USA
          Tel: 520-621-7982
          Fax: 520-621-1697
          Email: chm@u.arizona.edu
          Best blastx match: 'gb|A064314.1| (AC002335) hypothetical protein
          [Arabidopsis thal. . . 165 3e-40'. An open reading frame exists.
          Insert Length: 2 Std Error: 0.00.
FEATURES
source
1..701
location/Qualifiers
  /organism="Arabidopsis thaliana"
  /mol_type="mRNA"
  /ecotype="Columbia"
  /db_xref="taxon:3702"
  /tissue_type="leaves, flowering plants"
  /dev_stage="12 weeks"
  /clone_1ib="AA"
  /note="20 h 200mM NaCl"
ORIGIN
Alignment Scores:
Pred. No.:      2.28e-45      Length:      701
Score:          693.00      Matches:      140
Percent Similarity: 96.62%      Conservative: 5
Best Local Similarity: 94.59%      Mismatches:  3
Query Match:    14.43%      Indels:       2
DB:             2           Gaps:         0
US-09-920-705-3 (1-901) x BE037904 (1-701)

```

Qy 1 MetalaleuserteltyProphetaArgAlaaspSerGlyPheGlnSerAsnLeu 20
 Db 245 ATGGCGTTATCTTAAGAACCATTCAGAGCCATGATTCGGTTCCAGTCAAAACATCTT 304
 Qy 21 TrpValGlySerLeuThrProGluThrThrgluserAspleuThrGluLeuPheGlyArg 40
 Db 305 TGGGTCGGTAGCCTTAACCCGAGAGCAGACAGTCAGATCTGACCGAATGTTGGAAGA 364
 Qy 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 Db 365 TACGGCGATATTGATGATGATACCGGTGATCTTCACAGAGGCTTGCCTTTATATACAC 424
 Qy 61 ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 425 AGACATGTGAGAGAGACGATCGACCCAAAGAGCTCTTCAAGAGCAATTTGAATGA 484
 Qy 81 SerGlnIleValIleGluTyrAlaArgProAlaValProCysIleSerLeuTyrValGly 100
 Db 485 AGTCAAAATTAAAGATCAAGATCGACAGACCGCAAAACCTTGAAGATCTATGGGTGGT 544
 Qy 101 GlyIleGlyProAsnValSerIleAspAspLeuGluGluGluPheSerIlePheGlyLys 120
 Db 545 GGAATCGGCTCTTAATGTTCCAGAGATGACCTGACAGAGAGTTTCAGCAAGCTTGGGAAA 604
 Qy 121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
 Db 605 ATCAGAGATTTTAAGTCTTCAAGAACCGCAAGACGCTTTCATGCA-TATTATGAGATG 663
 Qy 141 AspAspAlaLeuGlnAlaValSer 148
 Db 664 GCTGATGC-TTACAGGCTAAGACA 686

RESULT 9

LOCUS AU237346 635 bp mRNA linear EST 01-APR-2002
 DEFINITION AU237346 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-16-A15 5',
 mRNA sequence.
 AU237346
 AU237346.1 GI:19876515

ACCESSION AU237346
 VERSION AU237346.1
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 635)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
 Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J.,
 Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A.,
 Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.

TITLE Large scale analysis of Arabidopsis full-length cDNA
 JOURNAL Unpublished (2002)
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4339
 Fax: 81-298-36-9060

REFERENCE Email: meeki@rtc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998) cDNA cleaved with BamHI
 and XhoI was ligated to modified Lambda FL-C1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified pBluescript vector. Please visit our web
 site (http://www.gen.riken.go.jp/e/Plant/index_e.html) for further
 details.

FEATURES
 SOURCE location/Qualifiers
 1..635
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL16-16-A15"
 /lab_host="DH10B"

ORIGIN /clone lib="RAFL16"
 /note="Site_1: BamHI; Site_2: SalI; dark-grown"

Alignment Scores:
 Pred. No.: 1,61e-38 Length: 635
 Score: 607.00 Matches: 123
 Percent Similarity: 97.62% Conservative: 0
 Best Local Similarity: 97.62% Mismatches: 2
 Query Match: 12.64% Indels: 2
 DB: 1 Gaps: 0

US-09-920-705-3 (1-901) x AU237346 (1-635)

Qy 1 MetalaleuserteltyProphetaArgAlaaspSerGlyPheGlnSerAsnLeu 20
 Db 255 ATGGCGTTATCTTAAGAACCATTCAGAGCCATGATTCGGTTCCAGTCAAAACATCTT 314
 Qy 21 TrpValGlySerLeuThrProGluThrThrgluserAspleuThrGluLeuPheGlyArg 40
 Db 315 TGGGTCGGTAGCCTTAACCCGAGAGCAGACAGTCAGATCTGACCGAATGTTGGAAGA 374
 Qy 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 Db 375 TACGGCGATATTGATGATGATACCGGTGATCTTCACAGAGGCTTGGCTTATATACAC 434
 Qy 61 ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 435 AGACATGTGAGAGAGACGATCGACCCAAAGAGCTCTTCAAGAGCAAAATTTGAATGA 494
 Qy 81 SerGlnIleValIleGluTyrAlaArgProAlaValProCysIleSerLeuTyrValGly 100
 Db 495 AGTCAAAATTAAAGATCAAGATCGACAGACCGCAAAACCTTGAAGATCTATGGGTGGG 554
 Qy 100 YGlyIleGlyProAsnValSerIleAspAspLeuGluGluGluPheSerIlePheGlyLys 120
 Db 555 TGGATCGGCGCTTAATGTTCCAGAGATGCGCTGAGAGAGAGTTTCAGCAAGTTTGGNAA 614
 Qy 120 IleGluAspPheArg 125
 Db 615 ATC-GAGATTTTAAAG 629

RESULT 10

LOCUS BH851246 416 bp DNA linear GSS 13-JUN-2002
 DEFINITION BH851246 SALK_072730.56.00.x Arabidopsis thaliana TDNA insertion line
 Arabidopsis thaliana genomic clone SALK_072730.56.00.x, genomic
 survey sequence.
 BH851246

ACCESSION BH851246
 VERSION BH851246.1
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 416)
 Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R.,
 Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
 Shim, P., Zimmerman, J. and Becker, J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the
 JOURNAL Arabidopsis Genome
 COMMENT Contact: Joseph R. Becker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379

REFERENCE Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within an annotated exon of At2g34340.
 Class: TDNA tagged.

REFERENCE	AUTHORS
1 (bases 1 to 817)	Kim, H., Yu, Y., Kudr...

TITLE	Global assembly of Cotton ESTs
JOURNAL	Unpublished (2004)
COMMENT	Contact: Rod A. Wing

FEATURES

Location/Qualifiers	Source
1. .817	

ORIGIN

Alignment Scores:

Pred. No.:	1,43e-32	Length:	81
Score:	535.00	Matches:	133
Percent Similarity:	57.60%	Conservative:	30
Best Local Similarity:	47.00%	Mismatches:	64
Query Match:	11.14%	Indels:	57
DB:	7	Gaps:	9

US-09-920-705-3 (1-901) X C0095058 (1-817)

QY	292	PhaTYSer1IwEtyIbArSArXgTIRApMePheAsnaAsp-----	307
Db	16	TTTGATTCAAGCATACATGAGAAAAGGCGTCGACGTCTTTGGACCATCTATTAGG	75
QY	308	-----ProSerCyValSer--	312
Db	76	CCTTCACAAATGACATCTTGCTCAAAATATCCATTACTACCAAAACGGTCTCGA	131
QY	313	SePProh1SerThngIyIleProGlySerMeArGProLeuArg-----	322
Db	136	TCATTACCATCTAGTGGCATTTTTGGATCAAAATGCCATAGCTCTTTAAGTCATCA	199
QY	328	GIYThzAsnGluArGSeTyIraEnGIyAlaGIuTyAsnAspValGIyLysGluPro	347
Db	196	GGTAGCTATGATCTTGTGCTTCAGGTCAAGATPACATGATGAT--TGGTCGACAGCAT	252
QY	348	AsnTPArArGrProSerAlaAsnGIYThngIyIleuProSerProThngIyProGly	366
Db	253	AATCGGGAAGGCTATCT-----CGTGTCTTCTGCTCAAGT	293
QY	368	IleuArProSerProAlaGInGIYThrArGrProMeArGrSerAsnProAspSerTrp	387
Db	292	TTTAGGCTTCCACGAGACAGCATCT-----GGTTCATGG	322
QY	388	GIuGIyTyArSpProAlaGInLeuValArGJusErIyArGIYThrArGrAspGIySer	407
Db	328	GATGTTTATACCTGAACCAATTTCAGAGAGATGACAAACGTTGACGATAGAGGCTGCC	387
QY	408	-----ValAspGIyPheTrpPromet-----GIYValAspGIuArg	413
Db	388	ATGGCTGTATATGATCTTGACCAAGTAATTGGGGGGCTTCAGAGTCCATTGTGAT	444
QY	420	SePheGluArgGIySerVal-----AlaAlaArgProIle	433

[illegible]

RESULT 1:

CA766318
100716

DEFINITION

to unknown, mRNA sequence.

VERSION CA766318.2 GI:27548346

SOURCE

100

REFERENCES

TTTT.E

JOURNAL
COMMENT

FEATURES

were collected from control (well watered) and stressed

ORIGIN

plants at 2 days before heading, at heading, 50% flowering
and 4 days after 50% flowering."

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 532.50	800	126
Percent Similarity: 55.51%	Conservative: 25	
Best Local Similarity: 46.32%	Mismatches: 84	
Query Match: 11.09%	Indels: 38	
DB: 6	Gaps: 7	

US-09-920-705-3 (1-901) x CA766318 (1-800)

```

QY 363 ProthrglyProgllyileLeuProSerProalaglnGlyThrArgaArgProMetArgSer 382
DB 8 CCTGTTCTGGTGTATTTCTTCATCAGCGAATCTGA-----AATCTCTCTTACT 61
QY 383 AsnProaspSerTrpGlnGlyTyrAspProalaglnLeuValArgGlnSerlySargThr 402
DB 62 GAGAGATCTGCTCAGAGTACATGATCA-----AGAGAGCGCAGAGAGTTG 109
QY 403 ArgArgSerArgSerValaap-----GlyPheThrPro 413
DB 110 CGTTGATGCTGCTCAGCTGACCTTATGATACAGAGGAGTGTAGAGGCTCTTATTCT 169
QY 414 MetGlyValaAspGlnaArgSerPheGlyArgGlySerValaAlaAlaArgPro----- 430
DB 170 GCTGGGATCTCTCAGCCGAGAGTGCAGCTTTCAGAGGAGCTCAAGTCTGCTATTCCA 229
QY 431 IleArgGlyProProaspSer-----AspHisIleTrpArgGlyMetIleAla 446
DB 230 ATCCATGCGACAGTGCACCAACATCATACCTTGAGCACTTGCGCTGTGATGATGCC 289
QY 447 LysGlyGlyThrProValaCysGlyAlaArgGlyValaProMetGlyLysGlyIleGlnThr 466
DB 290 AAAGGTGATCTCTGATGTCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 349
QY 467 LysLeuProGlnValaAlaAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHis 486
DB 350 CCTTACCGAGATGTTGTTATTTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 409
QY 487 TyrAlaValaAlaIleGlyCysGlnIleValaPhePheValaProaspArgGlnIleAspPhe 506
DB 410 TATAGATGCTCTCAGGGTGTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 469
QY 507 AlaserYrthrGlnPheLeuArgTyrLeuSerSerlySargAlaGlnValaAlaLys 526
DB 470 GTTCTTACACTGATATTTTGGCTACTTACGCTCAAAAAGTGGGCGAGGCTGTGAAG 529
QY 527 LeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspValaLeu 546
DB 530 GNTGATGGGGGACTACTTATTTTGTGCTCCACCTTCCGATTTTTCAGAAATGCTTGG 589
QY 547 GlnValThrArgGlnGlnaArgLeuTyrGlyValaValleuLys----- 560
DB 590 CAAGNTGATGTCAGAAAGCTCTTACGCTGATGATTAATATTCGCAATGCTGCTG 649
QY 561 -----LeuProProAlaValaProValThrLaserYrArgGlnGln 575
DB 650 CTGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 709
QY 576 SerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnLaser 595
DB 710 TCACAACT-----TTCTACG-ATGNAAGAGGAATTCCTTTGCGAGAAAG 756
QY 596 HisSerLeuTyrProProaspGlnaArgTyrIleArg 607
DB 757 GATAGTATGATTACNCTAGCAACAATACCATCGA 792
LOCUS CK208950 1105 bp mRNA linear EST 08-DEC-2003

```

DEFINITION

FGAS020677 Triticum aestivum FGAS: Library 5 GATE 7 Triticum

aestivum cDNA, mRNA sequence.

ACCESSION

CK208950.1 GI:39571340

VERSION

EST.

KEYWORDS

Triticum aestivum (bread wheat)

SOURCE

Triticum aestivum

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

1 (bases 1 to 1105)

Allard, F., Crosby, W. L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,

Gensewein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroche, A.,

Penkner, C., Roach, J. L. and Sahar, F. Muzak, I., Nilson, D.,

Functional Genomics of Abiotic Stress in Wheat and Canola Crops

Unpublished (2003)

Contact: Wm L Crosby

Bioinformatics

University of Saskatchewan, Department of Computer Science

1C101 Engineering Building, 57 Campus Drive, Saskatoon,

Saskatchewan, S7N 5A9, Canada

Tel: 306 966 1769

Fax: 306 966 2033

Email: fgas.estecs.usask.ca

This sequence is the direct result of the base calling software

Phred (default parameters). It is the raw base calls. To aid in the

identification of the high quality insert the software Lucy

(default parameters) has been run on this sequence. Lucy identified

the region (13,760).

Plate: L5B015 row: N column: 21.

Location/Qualifiers

1..1105

/organism="Triticum aestivum"

/mol_type="mRNA"

/db_xref="taxon:4565"

/clone_1ib="Triticum aestivum FGAS: Library 5 GATE 7"

/note="Vector: PCMV.SPORT6; Crown and developmental stages

of spike formation in wheat cultivar Norstar. 4 mRNA

populations were combined before constructing the library.

The first mRNA population is from 1cm crown sections after

30 days of cold acclimation. The second is from 1cm crown

sections after 11 days of deacclimation (before

deacclimation plants were fully vernalized for 49 days).

The third is from different developmental stages of spike

formation (5 to 50mm) that still have not emerged from the

leaf (dissection required). The last is from different

developmental stages of spike and seed formation after

having emerged from the leaf (visible). First strand

synthesis in this library was done in the presence of

methylated dCTP thereby protecting from internal cleavage

with NciI."

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 532.50	1105	152
Percent Similarity: 50.88%	Conservative: 50	
Best Local Similarity: 38.28%	Mismatches: 97	
Query Match: 11.09%	Indels: 99	
DB: 7	Gaps: 18	

US-09-920-705-3 (1-901) x CK208950 (1-1105)

```

QY 439 HisIleTPArgGlyMetIleAlaLysGlyThrProValaCysAlaArgCysVal 458
DB 87 CACTTCTGCGCGGAGAGTATGCGAAGGTGATCTCTGTTTGTGCGTCTGCTGCTG 146
QY 459 ProMetGlyLysGlyIleGlnThrLysLeuProGlnValaAlaCysSerAlaArgThr 478
DB 147 CCTATAGAGAGGCGATGATACCTTTACCGGACATGTTGTAATATCTCCGCTAGACA 206
QY 479 AspLeuAsnMetLeuAlaLysHisTyrAlaValaAlaIleGlyCysGlnIleValaPhePhe 498

```

```

Db 207 GATTGATATGCTGGAAAACACTAGAGATGCTCGGGTTTGAAATGCTTCTTC 266
Qy 499 ValProAspArgLugluAspPheAlaSerThrGluPheLeuArgTyrLeuSerSer 518
Db 267 TTACCAAGACAGTGAAGATGATTTGTTGCTTACCTGAATTTCTGCGGCTACTGGGCTCA 326
Qy 519 LysAspArgLagLValAlaLysLeuAspArgLysThrThrLeuPheLeuValProPro 538
Db 327 AAAGTGGCGCAGGGGTGTTAAGTTGAAGCAGAGAACCACTGTTGTTGGTCCGCCCA 386
Qy 539 SerAspPheLeuThrAspValLeuGlnValThrArgGlnLysLeuArgTyrGlyValVal 558
Db 387 TCGGATTTCTCGACGATGCTTTCGACGATGAGTGTCCAAAGCCCTTTATGGCGTGTGA 446
Qy 559 LeuLysLeuProPro-----ProAlaValProValThrAlaSerTyrArgGlnLys 575
Db 447 CTGCACATCTCCACGATGCTCAATGCTGCTCCAGAGACCTGACTGACGACGAG 506
Qy 576 SerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSer 595
Db 507 TCACAAAGCT-----TATTATGAT-----GGTGGT 530
Qy 596 HisSerLeuTyrProProArgGluAsnTyrIleArgGlyAlaPro-----GluHis 612
Db 531 GATACCATGCTAGCGATGCAAGAAATTACATATGAGCTCTGCTAATGGCAACCATCAT 590
Qy 613 LeuThrAlaAlaSerLysProSerValSerGlu----- 623
Db 591 CAGAGATCTGATTATCGGGGCTCTTTCGTGAGAGACAGTTCAGTCAAGGTATCAAGC 650
Qy 624 ---ProLeuArgLysProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGlu 642
Db 651 TTCCCATG-----ACCCAAATGACGAGACAGACAGACACACCTCAAGCCCTGAT 701
Qy 643 LeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSer 662
Db 702 ATTATGGCCCATTTAGCCAAACTTATATGCA-----AATGTCATGATCATGTTCCAGCAAT 758
Qy 663 HisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeu 682
Db 759 AGTCAG-----GTGGGTAACTTTCAGCAATCCCGTCAGACAG 794
Qy 683 TyrAsnGlyLysLysProSerGln-----AlaTyrLysArgGlyPro 696
Db 795 TTGGGACGAGCAAGCTCCACACGCTATGAGGGGCAATGCTGTGTCAGAGCAGACCA 854
Qy 697 GlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrPro 716
Db 855 -----ATACATCTGCT-----TACAAACCTGAGGTTACCTTA 887
Qy 717 -AlaGlyGlnLeuProProProSerArgTyrProProAlaSerAsnAsnProAsnTyr 736
Db 888 AGCT-----TTGGCTCCACAACCT-----CCTCTCCCTTAAAAAGAACCCNGT 932
Qy 736 rThrSerGlyMetValHisGlyAsnMetGlnTyrGlnSer----- 749
Db 933 T-----TCTCTCCCTTACACACGATTTTAAACAGAAAACCTT 971
Qy 750 -----GlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAs 766
Db 972 TTACCCCTCTCAAAAAATTAAACAGTCCCTTATTAACAG--TTACCCAGCTTA----- 1021
Qy 766 AsnTyrSerMetTyrThrGlnLysSerSerAsnHisProValSerGlnProMetValG1 786
Db 1022 -----ATTGCCGCTCCAAAATAATACA 1041
Qy 786 nGlnTyrGlnProGluAlaSerMetPro-----AsnGlnAsnTyrGlyPro 801
Db 1042 ACCTTACCGGCCCC-----CTCCACACCAAAAAGAAATGATGGCCCC 1081

```

```

LOCUS CF322873 582 bp mRNA linear EST 18-AUG-2003
DEFINITION HDN--02-E23.g1 OshDACL-overexpressing transgenic rice lambda phage
CDNA library II (HDN) Oryza sativa (japonica cultivar-group) cDNA
clone HDN--02-E23, mRNA sequence.
ACCESSION CF322873
VERSION 1
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriatroidae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 582)
AUTHORS Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,
Song,S.I., Kim,J.K., Kim,Y.-K. and Naim,B.H.
TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
COMMENT Contact: Naim B.H.
Genomics and Genetics Institute, Greengene Biotech Inc.; Division
of Bioscience and Bioinformatics, Myongji University
Yongin, Kyonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnaem@gbio.com, bhnaem@bio.myongji.ac.kr.
FEATURES
source
location/Qualifiers
1..582
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDN--02-E23"
/cisue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OshDACL-overexpressing transgenic rice lambda
phage cDNA library II (HDN)"
/notes="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; cDNA was inserted into lambda Uni-ZAP XR vector at
5' end with EcoRI and 3' end with XhoI site. mRNA was
derived from rice Histone Deacetylase overexpression
line."
Alignment Scores:
Pred. No.: 6,71e-31 Length: 582
Score: 511.50 Matches: 107
Percent Similarity: 68.21% Conservative: 26
Best Local Similarity: 54.87% Mismatches: 49
Query Match: 10.65% Indels: 13
DB: 7 Gaps: 5
US-09-920-705-3 (1-901) x CF322873 (1-582)
Qy 438 AspHisIleThrArgLysMetIleAlaLysGlyThrProValCysCysAlaArgCys 457
Db 8 GAGCAGCTTCTGGGTGAGTACATTGCCAAGGAGTGAATCTCTGATGTCGCTCGCTGT 67
Qy 458 ValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArg 477
Db 68 TTCCCTATTAAGAGAGGAGGTGTGAGATACCTTTACCGGAGTGTGTTATGTTCCGCTAGA 127
Qy 478 ThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhe 497
Db 128 ACTGAGCTGATGATGCTGGCAAGCACTTTATAGACGCTTCAGAGGTTGATATCGCTTC 187
Qy 498 PheValProAspArgLugluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSer 517
Db 188 TTCTTGCAGATAGTGAAGATGACTTTGTTCTTACACTGAATTTTTCGCTACTTAGGC 247
Qy 518 SerLysAspArgLagLValAlaLysLeuAspArgLysThrThrLeuPheLeuValPro 537
Db 248 TCAAAAGTCGGCAGGGGCTTGAAGGTTGAAGGGGGGACATCTTATTTTGGTCCCA 307

```

